ESSAYS IN DEMOGRAPHIC ECONOMICS

by

Steffen Reinhold

A dissertation submitted to The Johns Hopkins University in conformity

with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

August, 2007

Abstract

In the first of the essays, I reassess the relationship between premarital cohabitation and marital instability both theoretically and empirically. It has become a stylized fact that premarital cohabitation is positively correlated with the likelihood of marital dissolution. This is counterintuitive because economists expect that couples learn about each other during cohabitation and would only get married if they anticipated a successful marriage. One prominent explanation for the antithetic empirical evidence is self-selection of individuals with lower prospects of successful marriages into premarital cohabitation. Using U.S. data from 1988 to 2002 and duration models, I demonstrate that the positive relationship between premarital cohabitation and marital instability has weakened over time, and that the two are no longer associated with each other. A strong decline in this association within the group of more educated women drives the result. I hypothesize that a decline in the benefits of marriage has led to greater cohabitation and hence less self-selection within this group. Causal modeling using matching and panel models uncovers a negative effect of cohabitation on marital instability.

In the second essay, I investigate the effect of teenage childbearing on high school completion, and why alternative sets of instruments result in differing coefficient estimates. The three main reasons for these discrepancies are defective instruments, treatment effect heterogeneity, and multiple mechanisms by which instruments affect the treatment. I use two instruments, age at menarche and the occurrence of a miscarriage, to investigate which of these is likely to hold. While I do not find significant treatment effect heterogeneity, I find some indications that there may be problems with the instruments' validity. Furthermore, miscarriage disproportionately affects very young teenagers which are in turn more likely to drop out of high school. This may explain differences in instrumental variable estimates if age at birth is an additional explanatory variable.

Keywords:	Cohabitation, Marriage, Divorce, Teenage Childbearing, Duration Models, Instrumental Variables, Treatment Effects
JEL Classification:	C14, C21, D83, J12, J13
Advisors:	Professor Robert A. Moffitt Professor Tiemen Woutersen

Acknowledgements

Along the way, many people helped me in completing this dissertation. I am grateful to my advisor Robert Moffitt for his valuable advice and guidance throughout. His thoughtful comments on my work were a great inspiration to me. I am also thankful to Tiemen Woutersen for his encouragement and help while completing the dissertation. Their enthusiasm for research provided me with an example I wanted to follow. I also would like to thank my committee members Matthew Shum, Andrew Cherlin, and Constantine Frangakis for their helpful comments.

My wife Melanie supported me while writing the dissertation, and I thank her for her understanding and keeping her good humor. Even though my family could only support me from afar, I always was able to count on their support when necessary. Finally, I would like to thank all my friends at the Johns Hopkins University and in Baltimore for sharing this experience with me, especially Alena Bicakova, Ethan Miller, Viplav Saini, Tomas Svitil, Kevin Thom, and my fellow Ph.D. graduates Silvio Contessi, Witold Czubala, Anubha Dhasmana, Chris Metcalf, and Migiwa Tanaka.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

At least since Gary Becker's seminal work on the economics of the family (Becker 73; Becker, Landes, and Michael 1977; Becker 1981) economists started applying economic theories to explain demographic behavior outside the market sector such as fertility and marriage. This approach has been controversial from the beginning, and critics pointed to the problems associated with, for example, treating children as analogous to consumer durables (Blake 1968) and thereby neglecting the social context and motives of parents. On the other hand, labor economists were open to this approach because many important economic questions cannot be adequately answered without addressing demographic behavior at the same time. For example, women's labor supply and fertility choices are clearly not independent from each other. The essays in this dissertation are in this tradition and address empirically the relationship between cohabitation, marriage, and the risk of divorce in one essay and the educational consequences of early fertility in another essay.

Common to both essays is the econometric problem of self-selection. Demographic behavior such as cohabitation or childbearing is an endogenous choice variable for the individuals. For example, if one tries to assess the socioeconomic consequences of adolescent childbearing, one has to take into account the fact that very young mothers are not a random sample of the population and that they may have based their fertility decisions on anticipated consequences. Labor economists have early on addressed problems of a similar nature. Roy (1951) recognized theoretically the importance of accounting for individuals' choices of their profession when discussing the distribution of earnings. In empirical labor economics, Heckman pioneered econometric model with self-selection and regressions with discrete endogenous choice variables (Heckman 78, 79) . As one of the early empirical applications, Lee (1978) studied the effect of union membership on wage rates. A positive relationship between unionization and wage rates could arise if union membership causes wages to rise or if more productive workers are more likely to be unionized. Angrist and Krueger (1999) provide an overview of the now sizeable literature on causal modeling based on the notion of counterfactuals. For example, one could ask what wage rate a unionized worker would earn if he was not unionized. The difference between the wage rate the unionized worker actually earns and the counterfactual is the causal effect of unionization on the wage rate for this individual. The challenge in this sort of analysis is that no individual is ever observed in the counterfactual state. Randomization as an experimental technique circumvents the problem of self-selection by randomly assigning individuals to a certain state. Hypothetically, if one randomly assigned union status to workers, one could simply estimate the causal effect of unionization by comparing the means of wages across the two groups. But since randomization of individuals is often not possible in a social science setting, other econometric methodologies have been developed, most prominently instrumental variable techniques, matching estimators, differences-in-differences, and panel models to identify and estimate causal effects. Instruments are variables that affect the outcome of interest only because of their impact on the endogenous choice variable, and they often have been interpreted as 'natural' experiments akin to randomization. 'Natural' experiments rely on exploiting situations where either government policies or the force of nature shift individuals from one state to the other. For example, Angrist and Krueger (1991) argue that the quarter of birth is related to educational attainment because of compulsory schooling laws and that the quarter of birth is otherwise not related to the individuals' characteristics. Moffitt (2005) discusses causal analysis in the context of population research.

In my first essay, I reassess the relationship between premarital cohabitation and marital instability. Cohabitation rates haven been rising in the United States and a substantial proportion of marriages is now preceded by premarital cohabitation which is defined as living together and having an intimate relationship without being formally married. From a theoretical standpoint, many economists initially believed that couples who have cohabited prior to their marriages would lead more stable marriages. They have learned about their future spouse and would only go through with their marriage if the marital prospects were promising. But earlier empirical work has found that this is not the case. Couples who have cohabited before marriage are more likely to get divorced. One potential explanation for this counterintuitive result is self-selection of individuals with lower marital prospects into cohabitation. In fact, it has been found that cohabitors came from lower socioeconomic strata and have had less attachment to the institution of marriage. Even though the individuals learn about each other during cohabitation and only marry if they deem the quality of the relationship high enough, it could be the case that self-selection is more important in determining marital instability, as the theoretical model of Brien, Lillard, and Stern (2006) suggests.

I reassess the relationship between premarital cohabitation and marital instability by using the three most recent cycles of the National Survey of Family Growth (NSFG). This allows me to study whether the relationship between premarital cohabitation and marital instability has been stable over time. Since cohabitation rates have risen considerably in the United States, one hypothesis is that the process of self-selection into premarital cohabitation has changed, and that differences between cohabitors and non-cohabitors in terms of the quality of their marital prospects have decreased. Investigating this hypothesis with data on first marriages only, I find that while cohabitors faced a significantly increased risk of marital dissolution in the NSFG 1988 dataset of around 40%, the differences in the probability of separation between the two groups has almost completely disappeared in the most recent NSFG 2002 survey. Further analysis suggests, that this change is driven by a decline in the association between premarital cohabitation and marital dissolution within the group of well educated women. At the same time, premarital cohabitation has changed its status among this group from a fringe phenomenon to a rather common practice, consistent with the view that self-selection is now not a severe issue in this group any more. The changes in the behavior of well educated women may be explained by a decrease in the benefits of marriage or an increase in the benefits of cohabitation.

In a next step, I model the causal effect that cohabitation has on marital instability. One approach is to assume that only observable characteristics determine the selection into premarital cohabitation, and then matching is a non-parametric way to capture the effect of cohabitation. The idea is to compare an individual who has cohabited before marriage with a non-cohabitor that is similar on all other observable characteristics to obtain a counterfactual outcome for the cohabitor. These results suggest a negative effect of premarital cohabitation on the premarital instability implying that cohabitation is not a risk factor for divorce in itself. Finally using data on an individual's multiple marriages, one can account for individual-specific effects which are constant across all marriages. Estimating the Lillard, Brien, and Waite (1995) model, I do find that there is now only insignificant self-selection of individuals with high divorce risks into premarital cohabitation and that cohabitation itself decreases the risk of marital dissolution. This new result is in contrast to the earlier results of Lillard, Brien, and Waite (1995) who found no causal effect of cohabitation and significant self-selection of high-risk individuals into premarital cohabitation. Finally, a novel fixed effects estimator (Lee 2003) uncovers that cohabitation reduces the risk of separation after accounting for person-specific effects.

The second essay addresses the relationship between teenage childbearing and high school completion. In US data one finds a close statistical association between teenage childbearing and poor socioeconomic outcomes for mothers and their children. However, it is not obvious how this relationship arises. Teenage childbearing could be a causal factor for poor future outcomes, and one potential mechanism is that early childbearing leads to higher drop out rates from high school which in turn negatively affects future earnings potential. On the other hand, teenage mothers are potentially a select group who would have had these poor outcomes even if they had not given birth early. Researchers employed different econometric methodologies for tackling this problem of self-selection. Geronimus and Korenman (1992) account for heterogeneity in the family background by comparing sisters who have timed their births at different ages. Their results suggests that self-selection may indeed be a problem and teenage mothers have unobservable characteristics which make them more likely to experience bad socioeconomic outcomes. Bronars and Grogger (1993) use another 'natural' experiment to assess the effects of teenage childbearing on future outcomes by comparing women who have given birth to twins to mothers who have had singletons. They argue that having twins amounts to having an unplanned birth and they find that this additional birth has negative effects on high school completion rates. However, it is very likely that the effect of having the first child is different from the effect of moving from having one child to twins, a fact the authors discuss in their study. Most relevant to my essay are the studies by Ribar (1994) and Hotz, McElroy, and Sanders (1997, 1999) who are using instrumental variable techniques and who treat teenage childbearing as a binary variable. Ribar (1994) uses three different instruments in a bivariate probit model. Two instruments, local abortion rates and the availability of obstetricians and gynecologists, are environmental variables, and the third instrument is the age at menarche. Presumably, teenagers who are younger at menarche are longer at risk for getting pregnant and therefore have higher teenage childbearing rates. He uses all three instruments jointly and does not find that teenage childbearing itself is a risk factor for dropping out of high school. However, if he only uses age at menarche as an instrument, he estimates that teenage childbearing has an even stronger adverse effect on high school completion compared to the estimates when treating teenage childbearing as an exogenous factor. Hotz et al. (1997, 1999) also use a binary indicator for teenage childbearing and the occurrence of a miscarriage as an instrument. Again, this has the notion of a 'natural' experiment. If miscarriages are truly random, one essentially compares teenagers who have become mothers with teenagers who would have become mothers had there been no miscarriage. Their instrumental variable estimates suggest that teenage childbearing essentially has no effect on high school completion. This wide range of estimates using instrumental variable techniques is somewhat unsatisfactory, and in my essay I put forward three main reasons of why instrumental variable estimates might result in a wide range and even contradictory results.

First, estimates using alternative sets of instruments could differ because there may be potential problems with the instruments themselves. Instruments which have low explanatory power for teenage childbearing could result in estimates that are even more biased than if one treats teenage childbearing as exogenous. If the instruments are invalid, they cannot be excluded from the main outcome equation. Second, there might be treatment effect heterogeneity in the sense that each individual has a different effect of teenage childbearing. If this is the case, the coefficient estimate on teenage childbearing using an instrument depends on the set of individuals the instrument moves from one state to the other. If two instruments affect two distinct set of individuals, the resulting estimates may be different even though the instruments are valid. Finally, there are different channels through which instruments affect the treatment and indirectly the outcome of interest, and these mechanisms might result in different effects attributed to different instruments. In this example, both age at menarche and miscarriage affect the probability of a teenage childbearing but in different ways. Miscarriages affect teenagers who have already been pregnant while age at menarche affects the timing of initiating first sexual intercourse and thereby indirectly the probability of getting pregnant. In an empirical exploration of this issue, I use the most recent National Survey of Family Growth (NSFG) and two instruments, age at menarche and miscarriage, to study the effect of teenage childbearing on high school completion. I focus on these two instruments because they have resulted in estimates which are not only quantitatively different but also qualitatively.

First, I estimate 2 Stage Least Squares (2SLS) models and assess whether the instruments are weak. F-tests of the significance of the instruments in the first stage regression reveal that the instruments do not seem to be weak, but there seems to be some problems with the instruments' validity. Having two instruments allows me to test overidentifying restrictions because for these tests one needs at least one more instrument than endogenous regressors.

Second, I estimate a model with treatment effect heterogeneity (Moffitt 2007). In this model, I uncover only insignificant treatment effect heterogeneity and conclude that treatment effect heterogeneity is not likely to be the reason for the discrepancies between instrumental variable estimates.

Third, I investigate more closely the mechanism by which the instruments affect the binary indicator for teenage childbearing by looking at the age distribution of teenagers at the beginning of their first pregnancy. Miscarriages disproportionately affect the youngest group of teenagers who are likely to have the potentially most adverse consequences of teenage childbearing while age at menarche has a more symmetric effect on teenagers in all age groups. These differences in the age distribution at first pregnancy are masked by using a binary indicator for teenage childbearing and may result in contradictory estimates when using alternative sets of instruments.

Chapter 2

Premarital Cohabitation and Marital Instability

Industrial countries have witnessed rising cohabitation[1] rates while at the same time first marriage and remarriage rates have declined (Bumpass and Sweet 1989; Bumpass, Sweet, and Cherlin 1991; Bumpass and Lu 2000). Economists are interested in cohabitation and marriage because the question of why individuals enter and leave committed relationships has large welfare implications both on the individual and societal level. At the same time welfare policies (Moffitt, Reville, and Winkler 1998) and tax policies may give individuals incentives to enter one form of relationship or the other. At the present, cohabitation is a common experience in the United States. In 2002, more than half of all women aged 19-44 have ever cohabited in their

[1]Cohabitation is understood here as living together under the same usual address and having an intimate sexual relationship.

lives. When cohabitation first emerged in the USA, it was mainly a phenomenon of the less educated and economically disadvantaged, but by now it has extended to the American middle class. While cohabitation is still more likely among the less educated, it has become a common experience even for the well-educated. Around half of all women with college or higher degrees have cohabited at least once as of 2002.

In this essay, I investigate the effect of these trends on the relationship between cohabitation and marital instability. Earlier empirical studies have found that marriages preceded by premarital cohabitation are less stable both for the United States (Booth and Johnson 1988; Teachman and Polonko 1990; DeMaris and Rao 1992) and Western Europe (Bennett, Blanc, and Bloom 1988). I show that this empirical regularity has broken down over the last twenty years, and I look for an explanation for this new finding both theoretically and empirically.

The essay is organized as follows: Using a theoretical search model of marriage and cohabitation developed by Brien, Lillard, and Stern (2006), I discuss the causal effect of cohabitation on marital stability and how the model generates the apparent positive relationship between cohabitation and subsequent marital instability. Couples learn about the quality of their relationship during cohabitation, and some of them decide not to go through with their marriage. On the other hand, cohabitors who get married can be more certain about their marriage. However, this effect is obscured by the self-selection of individuals with low match quality into premarital cohabitation. I add to their results by showing how a general decline in the benefits of marriage can lead to both more cohabitation and less self-selection.

The idea that couples learn about the match-specific quality during cohabitation goes back at least to Becker (1973) and Becker, Landes, and Michael (1977). Since cohabitors have a more precise estimate of their match quality, there should be fewer bad surprises during marriage. Based on this theoretical argument, one expects that former cohabitors lead more stable marriages. However, earlier empirical evidence pointed in the opposite direction. Self-selection is now an accepted explanation for these counterintuitive results (Schoen 1992; Lillard et al. 1995). According to this view of cohabitors as a select group, individuals who are at a higher risk of marital disruption also tend to cohabit before their marriage. This view is supported by the fact that cohabitors often have other elevated risk factors for marital disruption as for example lower education, unstable family background (Bumpass and Sweet 1989) and lower commitment to the institution of marriage (Bennett, Blanc, and Bloom 1988). To the extent that premarital cohabitation has become integrated in the regular courtship process, it may have become less signifying of individuals with elevated risk factors (Teachman 2002). As cohabitation has become more common there might be less self-selection on unobservables in the group of premarital cohabitors. But then the apparent positive relationship between premarital cohabitation and marital instability may weaken or even reverse its sign as the recent experience in Denmark (Svarer 2004) suggests. Furthermore, (Liefbroer and Dourleijn 2006) study 16 European countries and find that premarital cohabitation is associated with marital dissolution only in countries with either very high or very low rates of premarital cohabitation.

In countries where around half of all couples cohabit before marriage they don't find a statistical association between premarital cohabitation and marital instability. In their view, if only few people cohabit they are a select group. On the other hand, if a big majority cohabits before marriage, the couples getting married without prior cohabitation are also a select group. Premarital cohabitation has always been more common among the less educated while more educated women are still more likely to get married right away. Because premarital cohabitation was always more common for less educated, positive self-selection may not be a large problem within this group. However, for well-educated women, it has been rather uncommon to premaritally cohabit. For this reason, the self-selection on unobservables might have been different for the group of less educated compared to the well-educated.

In the empirical section of my essay, I use the three most recent cycles of the National Survey of Family Growth (NSFG) in 1988, 1995, and 2002 to study the evolution of the relationship between cohabitation and subsequent marital instability. Comparing observable characteristics of cohabitors and non-cohabitors across all three cycles of the NSFG, I show the evolution of risk factors for cohabitors and non-cohabitors. For all three cycles of the NSFG, I conduct proportional hazard regressions controlling for age, religion, race, educational achievement, fertility, and other socioeconomic variables to assess the relationship between the hazard of marital dissolution in the first marriage and premarital cohabitation. Since I use three cycles of the survey, I am able to determine whether and how the coefficient on premarital cohabitation has changed over time. Most other studies could only use a single survey and have not been able to study the stability of coefficients over time (an exception is Teachman 2003) . In addition, I estimate proportional hazard models with interactions between education and premarital cohabitation to see whether the coefficient on cohabitation has trended differently across educational groups.

Proportional hazard models cannot uncover a causal effect of premarital cohabitation on marital instability because the coefficient on premarital coefficient is tainted with self-selection. Unfortunately, there are no credible instruments correlated with the decision to cohabit but not with the error term in the marital dissolution process. For this reason, I use matching estimators and random and fixed effect models. Matching estimators invoke fewer parametric assumptions on the effect of cohabitation on marital stability. Furthermore, if there are substantial interactions between variables like education and the decision to cohabit, then the matching estimators may be better suited to deal with the interactions than the more restrictive proportional hazard models. Again, I compare the estimates for the three cycles of the NSFG to determine how these effects have changed over time. Random and fixed effects models are another potential way to deal with unobserved heterogeneity. Lillard, Brien, and Waite (1995) model the decision to cohabit and the marital dissolution process simultaneously with a random effects assumption and rely on the presence of multiple outcomes for one individual for identification. They find no causal effect of cohabitation on marital dissolution once the self-selection into cohabitation is taken into account. Furthermore, they find strong evidence that there is self-selection of high-risk individuals into cohabitation. I estimate the same model with a similar set of regressors using the 2002 cycle of the NSFG to study whether the process of self-selection has changed, allowing a direct comparison to the results in Lillard, Brien, and Waite (1995). Finally, I use a novel fixed-effect estimator (Lee 2003) as a way of accounting for the presence of person-specific heterogeneity that may be correlated with the decision to cohabit prior to marriage.

2.1 Theoretical Considerations

The Brien et al. (2006)[2] search model demonstrates that couples learn about their mutual compatibility during cohabitation, yet at the same time their future marriages are less stable because there is self-selection on marital 'quality' into premarital cohabitation. The decision to enter one living arrangement or the other depends on the underlying benefits of marriage and cohabitation. I show that the empirical observed decline of marriage rates and rise in divorce rates may be explained within their model by declining benefits to marriage. I also demonstrate within their model how a change in the underlying benefits of marriage affects the self-selection process on the underlying unobserved relationship quality.

Their theoretical search model is a partial-equilibrium model in which a woman receives matches from a stable distribution. The two-sided nature of the market and how the distribution of matches arises from the equilibrium is not modeled. The

[2] BLS henceforth.

relationship status of an agent is marked as

$$m_t = \begin{cases} 1 \text{ single} \\ 2 \text{ cohabiting} \\ 3 \text{ married} \end{cases} \tag{2.1}$$

The duration of the current relationship is denoted d_t. At the beginning of each relationship an unknown match quality, which is $\theta \sim N\left(0, \sigma_\theta^2\right)$, is drawn. The agent receives a signal of the underlying match quality as long as the relationship lasts. This signal is an AR(1)-process

$$\begin{aligned} \varepsilon_t &= \theta + \eta_t \text{ for } d_t = 1 && (2.2) \\ \varepsilon_t - \theta &= \rho\left(\varepsilon_{t-1} - \theta\right) + \eta_t \text{ for } d_t > 1 && (2.3) \\ \text{with } \eta_t &\sim N\left(0, \frac{\sigma_\eta^2}{1-\rho^2}\right) \text{ for } d_t = 1 && (2.4) \\ \text{and } \eta_t &\sim N\left(0, \sigma_\eta^2\right) \text{ for } d_t > 1 && (2.5) \end{aligned}$$

After observing the noisy signal, the agent updates her estimate of $\widehat{\theta}_{d_t}$ knowing all parameters of the process. This allows to treat $\widehat{\theta}_{d_t}$ and ε_t as state variables. It is also assumed that there is a time $\overline{t}_d$ after which no new information is revealed. The agent enjoys flow utility in each period from the current signal and a deterministic function of her characteristics, and her marital status.

The timing is as follows: If the agent is single, she gets a signal of a new match. She then has to decide whether to stay single, to start cohabiting, or to get married right away. The flow utility of being single is normalized to zero. If the agent is in a relationships, she receives flow utility from being either married, $f_t(3)$, or from cohabiting, $f_t(2)$. In general, these flow utilities will be a function of presence of children, the length of the relationship and her socioeconomic characteristics. These arguments are dropped for the moment. In addition, she receives an utility flow from the match quality, ε_t. These two components of the period utility are additively separable. In the cohabitation state, she can decide to split up, in which case she would have to pay the separation costs, D_2, and be single in the next period. Alternatively, she could continue cohabiting or marry her cohabiting partner. In the latter case, she would be married starting next period.

In the married state, the woman receives again per period utility. When married she can either continue being married or separate If she splits up, she will be single in the next period and have to pay the separation costs of D_3. The individual discounts the future with $\beta < 1$. Let $F(m_t)$ be the set of feasible choices for each marital status, then the agent's value function is:

$$\begin{aligned} V_t\left(m_t, m_{t-1}, \widehat{\theta}_{d_t}, \varepsilon_t\right) &= f_t(m_t) + 1(m_t > 1)\varepsilon_t - D_{m_{t-1}} 1(m_t = 1) \\ &+ \beta E_{\varepsilon_{t+1}}\left\{\max_{m_{t+1}\in F(m_t)} V_{t+1}\left(m_{t+1}, m_t, \widehat{\theta}_{d_{t+1}}, \varepsilon_{t+1}\right) \middle| \widehat{\theta}_{d_t}, \varepsilon_t\right\} \end{aligned} \tag{2.6}$$

BLS also assume that there is a point t^* after which no decisions can be done anymore and a point in time t^{**} after which the agent dies. This allows to solve iteratively for all $t < t^*$.

In their model, they make the following two assumptions:

$$D_3 > D_2 > D_1 = 0 \tag{2.7}$$

$$f_t(3) > f_t(2) \tag{2.8}$$

These assumptions are necessary for the coexistence of cohabitation and marriage in equilibrium. If $f_t(3) = f_t(2)$ then cohabitation dominates marriage and no one ever gets married. Also if $D_3 = D_2$ then marriage dominates cohabitation. The separation costs include all psychological and monetary costs of a separation including costs for using the legal system. Higher flow utility in marriage can be rationalized in several ways: marriage provides legal protection if the couple decides to have children, it allows couples to specialize, and perhaps there is greater social approval by other family members for this form of relationship. In a section below, I will discuss which factors might have affected changes in this $f_t(.)$-functions.

BLS show that there is a set of reservation values $\varepsilon_t^*(m_t, m_{t-1})$ governing the transition from state m_{t-1} to m_t. The value $\varepsilon_t^*(m_t, 1)$ ranks being single vis-a-vis co-residential unions, married or cohabiting. Formally, this reservation value is defined as

$$V_t\left(m_t, 1, \widehat{\theta}_t, \varepsilon_t\right) > V_t\left(1, 1, \widehat{\theta}_t, \varepsilon_t\right) \quad \forall \varepsilon_t > \varepsilon_t^*(m_t, 1) \tag{2.9}$$

$$V_t\left(m_t, 1, \widehat{\theta}_t, \varepsilon_t\right) < V_t\left(1, 1, \widehat{\theta}_t, \varepsilon_t\right) \quad \forall \varepsilon_t < \varepsilon_t^*(m_t, 1) \tag{2.10}$$

There is also a reservation value, $\varepsilon_t^*(1, m_{t-1})$ for women in a co-residential union determining whether they separate. This reservation value is defined as follows.

$$V_t\left(m_{t-1}, m_{t-1}, \widehat{\theta}_t, \varepsilon_t\right) > V_t\left(1, m_{t-1}, \widehat{\theta}_t, \varepsilon_t\right) \quad \forall \varepsilon_t > \varepsilon_t^*(1, m_{t-1}) \tag{2.11}$$

$$V_t\left(m_{t-1}, m_{t-1}, \widehat{\theta}_t, \varepsilon_t\right) < V_t\left(1, m_{t-1}, \widehat{\theta}_t, \varepsilon_t\right) \quad \forall \varepsilon_t < \varepsilon_t^*(1, m_{t-1}) \tag{2.12}$$

Because of the difference in divorce costs, they show that

$$\varepsilon_t^*(1,3) < \varepsilon_t^*(1,2) \tag{2.13}$$

implying that $\frac{\partial V_t\left(3,m_{t-1},\widehat{\theta}_{d_t},\varepsilon_t\right)}{\partial \varepsilon_t} > \frac{\partial V_t\left(2,m_{t-1},\widehat{\theta}_{d_t},\varepsilon_t\right)}{\partial \varepsilon_t} > \frac{\partial V_t\left(1,m_{t-1},\widehat{\theta}_{d_t},\varepsilon_t\right)}{\partial \varepsilon_t} = 0$

That is, women in marriages tolerate a lower match quality in comparison to cohabitors before they separate because of the difference in separation costs.

There is also a reservation value $\varepsilon_t^{**}(m_t, m_{t-1})$ determining whether to marry:

$$V_t\left(3, m_{t-1}, \widehat{\theta}_t, \varepsilon_t\right) > \max_{m<3} V_t\left(m, m_{t-1}, \widehat{\theta}_t, \varepsilon_t\right) \quad \forall \varepsilon_t > \varepsilon_t^{**}(3, m_{t-1}) \tag{2.14}$$

$$V_t\left(3, m_{t-1}, \widehat{\theta}_t, \varepsilon_t\right) < \max_{m<3} V_t\left(m, m_{t-1}, \widehat{\theta}_t, \varepsilon_t\right) \quad \forall \varepsilon_t < \varepsilon_t^{**}(3, m_{t-1}) \tag{2.15}$$

Of particular interest here, is the reservation value, $\varepsilon_t^{**}(3,2)$, determining whether a cohabiting woman will enter a marriage. BLS show that $\varepsilon_t^{**}(3,2) > \varepsilon_t^*(1,2)$. They also show that the value functions are continuous in the flow utility functions and if $f_t(3) = f_t(2)$, it follows that $\varepsilon_t^{**}(3,2) = \infty$. That is, if there is no utility bonus for being married, the reservation value for the decision to get married is infinite with the consequence that no one marries.

The Causal Effect of Premarital Cohabitation on Marital Instability

One possibility to define the causal effect of premarital cohabitation on marital instability is to define it on an individual level and to condition on the unobserved match quality. Conditional on the unobserved match quality, the effect of premarital cohabitation on marital outcomes is the change in the separation probabilities if the person cohabits and marries after a while versus if she immediately marries. Since I condition on observing at least one marriage, this definition of the causal effect corresponds to the causal effect I want to uncover in my empirical work. This effect should be negative according to this model.

Cohabitors who get married must have had relatively positive surprises concerning their match quality during cohabitation. Because the reservation values to separate, $\varepsilon_t^*(1,3) < \varepsilon_t^*(1,2)$, are higher for cohabitors, cohabitors who later get married are more likely to have experienced positive transitory shocks during their cohabitation than couples who get married immediately. If cohabitors had 'bad' transitory surprises in the beginning of their cohabitation, they could have separated. For this reason they will have a higher estimate of their match quality, $\widehat{\theta}$, as compared to couples who get married right away. Since divorce probabilities decrease in $\widehat{\theta}$, married couples who have cohabited have more stable marriages conditioning on the true relationship quality θ. The estimate of the true match quality is auto-correlated so that this positive effect only slowly wears off as cohabitors realize that they overestimated their match quality. One complication with this is that the reservation values while married for separating may be different depending on whether one has cohabited or not. However, if they are not too different then one should see that marriages preceded by cohabitation are more stable conditional on unobserved match quality. BLS analyze the case where a cohabitation precedes all marriages. Under this assumption, cohabitation decreases average marital instability on the population level by weeding out bad matches in the cohabitation phase.

Even though the effect of cohabitation on marital instability may be negative on the individual and population level there is a self-selection effect into premarital cohabitation. BLS show that married couples who cohabited have, on average, a lower match quality, obscuring the negative effect of premarital cohabitation on marital instability. Overall, this selection effect dominates and marriages preceded by premarital cohabitation are less stable which is consistent with earlier empirical evidence.

From this discussion, it is clear that empirical studies, that do not control for the unobserved match quality, will deliver biased estimates of the causal effect of cohabitation on marital outcomes. One further implication of the model is that the observed association between cohabitation and marital dissolution is the result of the causal effect of cohabitation and the self-selection of women with lower prospects of marital success into premarital cohabitation. In the empirical section of my analysis, I observe a decline in the correlation between cohabitation and marital instability over time. This decline can in principle be attributed to either a change in the causal effect or a change in self-selection. In the following section, I argue that the self-selection process has changed. The underlying benefits of marriage and cohabitation determine not only the incentives to enter marriage and cohabitation but also the selection on unobservable match quality.

Benefits to Marriage and Change in Self-Selection

In the following, I demonstrate that the BLS model implies that a change in the underlying benefits to marriage not only affects the proportion of women who cohabit but also determines the self-selection on unobservables. If one observes a large change in the proportion of cohabitors and married people in a population, one should suspect changes in the self-selection on unobserved match quality.

For the special case of constant benefits to marriages, $f(3)$, BLS show analytically that a decrease in $f(3)$ increases the hazard of divorce. The intuition is that if the benefits of marriage decline, there is less of an incentive to stay in a marriage. BLS conducted simulations with parameters estimated for US data asking what happens if the marriage credit is expanded, i.e.$f(3)$ increases. For their particular parameter estimates, they found that higher benefits of marriage increase the likelihood of cohabitors to both split up and to marry. For singles, it increases the likelihood of getting married but it decreases the likelihood of starting a cohabitation. Thus, the overall effect of an increase of the benefits to marriage is an increase in the proportion of married people and a decrease in the proportion of cohabiting women. Based on these results, a decline in the benefits of marriage could explain parsimoniously the observed rise in cohabitation, the decline in marriage, and the increase in divorce rates.

I add to these results by showing how a change in the benefits to marriage affects the reservation value $\varepsilon_{t+1}^{**}(3,2)$ governing the transition of cohabiting women into marriage. BLS already have shown that if $f_t(3) = f_t(2)$, it follows that $\varepsilon_t^{**}(3,2) = \infty$. Intuitively, if the benefits of marriage decline relative to the benefits of cohabitation, cohabitors who want to get married require a higher match quality to make this decision. The benefits of marriage are the higher flows during marriage while the disadvantage is the higher cost of separation. To offset the decrease in flow utility couples require a higher match quality which make the outcome of a separation less likely. I show in the appendix that $\varepsilon_{t+1}^{**}(3,2)$ is a strictly decreasing function in $f_t(3)$, and a strictly increasing function in $f_t(2)$.

Claim. $\varepsilon_{t+1}^{**}(3,2)$ is a strictly decreasing function in $f_t(3)$ and a strictly increasing function in $f_t(2)$.

Proof. See appendix. □

Thus, if the benefits to marriage decline, the average match quality of married women who have cohabited improves. A decline in the benefits of marriage also affects the incentives to marry immediately, and the average match quality of singles getting married immediately also improves. In general, it will be hard to say theoretically whether the average match quality improves stronger for cohabitors or for singles who get married. Even though there is some ambiguity about the overall effect, this demonstrates that self-selection on unobservables is a function of underlying benefits of marriage. If one observes large changes in marriage and cohabitation rates, one suspects that underlying benefits of marriage and cohabitation must have changed. But then, self-selection on unobservables must also have changed.

Declining Benefits of Marriage

In the following, I argue that benefits of marriage have declined in the US for all women. There is a wide range of explanations for the decline in the benefits of marriage. Most of them are not mutually exclusive but rather reinforce each other. But there are also important differences for the reasons of this decline for different socioeconomic groups. Some factors leading to a decline in the benefits to marriage may also affect the benefits to cohabitation. I will address this issue by using results by Song (2001).

Becker (1973, 1981) proposes an economic model of marriage. In this model, the incentive to marry stems from the possibility to divide labor and to specialize on activities where one is more productive than the spouse. One implication of this model is that the gains to marriage are higher in a situation where the pay differential between males and females is higher. Moffitt (2000) finds that marriage rates have gone down for all educational groups, but especially strong for the less educated. Within the Becker model this may be explained by rising female wage rates for more educated women and stagnating wage rates for less educated men. There are other economic reasons of why marriage rates may have fallen in the US. The welfare system might encourage women not to marry and to cohabit instead. Moffitt, Reville, and Winkler (1998) found modest evidence for this claim. Cherlin (1991) discusses cohort-based explanations of patterns of family formation. Cohort size partly determines economic opportunities later in life and thereby household formation.

Lichter, MacLaughlin, Kephart, and Landry (1992) proposes a 'shortage of marriageable men' for some women as a reason of a decline in marriage rates. This could affect less educated and African-American women stronger because of stagnating real wages for blue-collar workers and high incarceration rates among African-Americans.

Cherlin (1991) also discusses changing attitudes and values as a possible explanation. While the 1950's were more family-oriented, there was a cultural change towards more individualistic values. At the same time, women adopted less traditional gender roles. Amato and Booth (1995) have shown that if wives adopt non-traditional gender roles their perceived marital quality declines. Thus, a change in gender role attitudes may cause a decline in the perceived benefits to marriage for women. Cherlin (2004) argues that marriage in the United States has undergone a deinstitutionalization process, that is the social norms governing expectations of behavior in marriages have weakened. Husband and wife therefore have to negotiate what to mutually expect from each other, which is a potential source of conflict.

One might wonder whether some of the factors affecting benefits to marriage also have affected benefits to cohabitation. I argue here that there is some reason to believe that marriage and cohabitation are affected asymmetrically. One good example is public assistance which affects cohabitation and marriages asymmetrically as Moffitt, Reville, and Winkler (1998) demonstrates. Song (2001) investigates labor supply and fertility patterns in marriage and cohabitation. She found that labor supply for women is higher among cohabiting women than among married women. Thus, rising female wages for educated women might have an asymmetric effect on these living arrangements.

Fixed Effects in the BLS Model

Another extension of this model may rationalize the use of fixed effects model. In the BLS model, the functions $f_t(.)$ are deterministic functions of socioeconomic and demographic characteristics. However, one could also think of a model where there are personal unobserved characteristics that shift the benefits to marriage and cohabitation. Some of them may be fixed while others may be changing over time. Some of these unobserved components may also be related to other variables in the $f_t(.)$ functions. For example, a person with good education might also have developed better interpersonal communication skills, thereby increasing her benefits of being in a relationship. If these unobserved components are permanent, then fixed effects techniques may be preferable. To the knowledge of the author, none of the previous studies on marital stability and cohabitation uses a fixed effects approach. Lillard, Brien, and Waite (1995) take one step into this direction by using a parametric random effects model to account for these permanent effects. Instead, I use both random and fixed effects model to deal with this type of heterogeneity. Notice that this heterogeneity is different than in BLS where there is only match-specific unobserved heterogeneity but no person-specific unobserved heterogeneity.

2.2 Data

2.2.1 Description of the NSFG data set

The National Survey of Family Growth (NSFG) was conducted by the National Center of Health Statistics (NCHS) as a representative sample of women aged 15-44 for the years 1973, 1976, 1988, 1995, and 2002. Their main purpose is to provide information on marriages, divorces, fertility, and the health status of women and their children. The survey includes information on important events such as marriages and child-births along with other socioeconomic and demographic information. The survey asks retrospective questions for the full history of marriages and divorces; but only starting in 1988 it also included more detailed information on women's cohabitation history. Since I am interested in the effect of cohabitation on marital outcomes, women who never married are omitted. I analyze first marriages and cohabitation that preceded them. This left me with 5030 first marriages for 1988, 6776 first marriages for 1995, and 4030 first marriages for 2002. Because I condition on observing a first marriage, these samples include women aged 16-44 in 1988, aged 15-44 in 1995, and aged 17-44 in 2002.

2.2.2 Prevalence of Cohabitation

There is now ample evidence that cohabitation rates have risen in the past (Bumpass and Sweet 1989; Bumpass, Sweet, and Cherlin 1991; Bumpass and Lu 2000). Cohabitation has by now become a common experience among women in the

United States. Besides concentrating on premarital cohabitation, there are at least two ways to measure this rise in cohabitation rates. Bumpass and Lu (2000) use both the National Survey of Families and Households (NSFH 1987/1988) and the NSFG 1995 to calculate percentages of women ever cohabiting and currently cohabiting. I updated their data with new results from the NSFG 2002 and show these combined in tables 2.1 and 2.2.

In table 2.1, cohabitation rates by age group are shown. The first three columns show the percentage of ever cohabiting women by age group in 1987, 1995, and 2002. Overall the percentage of ever cohabiting women has risen steadily from a third in 1987 to well over half of all women in 2002. This rise in cohabitation rates was most marked for women aged 35-44. While in this group prevalence of cohabitation was below average 1987, it is now well above average. Furthermore, while cohabitation was more common among younger women in earlier years, older women are now very likely to experience cohabitation. The last three columns of table 2.1 show rates of currently cohabiting women (of not currently married) for the same years by age group. While in 1987 younger women aged 25-29 were the most likely to cohabit, the rate of currently cohabiting women rose quicker for the older age groups, closing the differences by age in cohabitation rates. The percentage of currently cohabiting women almost doubled for women aged 35-39 while the growth was more modest for other age groups. In empirical studies, it has often been found that a young age increases the risk of union disruption. The rise of cohabitation among older women would therefore be one additional factor stabilizing the relationships of cohabitors.

Table 2.2 shows percentages of ever cohabiting women by education and race for 1987, 1995, and 2002. While cohabitation is still more common among the less educated, cohabitation ceases to be a fringe phenomenon among the well-educated. Among highly educated women with a college degree, almost half have ever cohabited. Bumpass and Lu (2000) conclude that economic constraints could not explain the differentials in cohabitation rates among the different educational groups since cohabitation is so common for all groups. Table 2.2 also shows rates of ever cohabiting women by race. In all years, cohabitation is most common among blacks but the racial divide in prevalence of cohabitation has been slightly reduced. The strongest increase in cohabitation rates was among whites, continuing the trend identified in Bumpass and Lu (2000). To the extent that economic disadvantages are still associated with race in the United States this trend supports the argument that cohabitation has now extended to the middle class.

Table 2.3 shows the means of selected variables for women who cohabited before their first marriage and for women who did not.[3] The dramatic rise in premarital cohabitation mirrors the trends identified for ever cohabiting and currently cohabiting women By now, almost half of all first marriages are preceded by premarital cohabitation –up from about a quarter in 1988. The group of cohabitors has lower educational achievement than non-cohabitors in all years. At the same time the educational achievement improved for both groups reflecting the general trend of rising

[3]In the following, I refer to women who cohabited before their marriage as 'cohabitors.'

education levels in the United States. Cohabitors are on average younger on the day of the interview reflecting a cohort effect. Cohabitors are older at the day of their marriage, partly reflecting the time spent in cohabitation before marriage. It has been shown that young age is a predictor of marital dissolution, giving cohabitors a potential advantage. However, at the same time the age-difference between spouses is bigger for cohabitors which is a potential risk factor. There is an important difference in fertility behavior between cohabitors and non-cohabitors. Cohabitors are much more likely to have children outside of marriage. For both groups there is a large differential between premarital conception and a premarital lifebirth. One explanation for this is that women marry after conception, possibly to legitimize the child.

Table 2.4 shows cohabitation status relative to the first marriage for all educational groups. In addition, it shows the percentage of intact marriages at the time of the interview by education and cohabitation status. Married women with less than a high school education have the highest rates of premarital cohabitation in all three cycles of the NSFG. For all educational groups, one observes a strong rise in premarital cohabitation. Even among women with some education, more than forty percent of women have cohabited with their spouse. In general, marriages of well-educated women are more likely to be intact at the day of the interview. However, well-educated women also have a shorter time at risk since educated women marry later.

The use of a binary indicator for premarital cohabitation may not be completely adequate and has recently been criticized by sociologists because it may hide some important qualitative differences (Manning and Smock 2005). For example, a formal engagement before cohabitation with a clear understanding that a marriage is planned might change the expectations and behavior of the couple during this phase. In the NSFG 2002 women were asked whether they were engaged while cohabiting. I find that engagement and cohabitation combined increases marital stability in the NSFG 2002. I do not use this measure further because this particular question was not asked in earlier cycles of the NSFG making it impossible to study this effect over time. Similarly, one might worry that the average length of premarital cohabitation is important in determining its effect on marital stability. For this reason, I examined whether a cohabitation shorter than three months has a different effect than longer cohabitation, but I did not find differences in the coefficients on these two measures of cohabitation. Also, I do not find that cohort effects were important in determining whether marriage was preceded by a long or short cohabitation. My justification to use the binary measure is that the positive effect of cohabitation on marital instability has been found in many different datasets in which cohabitation was measured differently. Thus, the empirical relationship is robust to the exact definition of cohabitation in the particular data set, and my definition of cohabitation is similar to those used in the literature.

2.3 Empirical Models and Results

2.3.1 Proportional Hazard Models

In a first step, I use proportional hazard regressions to facilitate comparison with earlier empirical work. The proportional hazard model was introduced in Cox (1972). The hazard in the simple proportional hazard model can be written as follows:

$$h\left(t|X\left(t\right)\right)=h_0\left(t\right)*\exp(X\left(t\right)'\beta) \tag{2.16}$$

That is, the hazard at each point in time factors into two components, one that only depends on time $(h_0\left(t\right))$, the other only depends on the value of the covariates, $\exp(X\left(t\right)'\beta)$. The proportional hazard model is semi-parametric and the baseline hazard $(h_0\left(t\right))$ does not need to be specified but is estimated non-parametrically. Furthermore, notice that there is no unobserved heterogeneity in this specification. Therefore, the coefficient on cohabitation incorporates both any causal effect of cohabitation on marital duration and possibly the self-selection of high risk individuals into cohabitation.

In table 2.5, I present proportional hazard regressions for the 1988, 1995, and 2002 cycles of the NSFG. The sample consists of first marriages of women who are younger than 44 in each year. No marriages of women younger than 15 years old are observed. The dependent variable is the hazard of marital dissolution for the first marriage. All coefficients are reported as hazard ratios: a coefficient of greater than one indicates that this regressor increases the risk of marital dissolution while a coefficient smaller than one indicates a decrease in risk. I chose other explanatory variables that were found as predictors of marital success in earlier studies. These include education, race, religion, fertility indicators, and age at marriage.

For each of the cycles, I estimate the model including a dummy variable for premarital cohabitation and a different model where the cohabitation dummy is interacted with education. In the specification with a dummy for premarital cohabitation, one observes a marked decline in the coefficient on premarital cohabitation. The hazard of marital dissolution for cohabitors is about 40% higher than for non-cohabitors for the NSFG 1988. By 1995, the hazard of marital dissolution is about 27% higher for cohabitors, and the effect becomes even smaller in 2002, where cohabitation statistically no longer affects marital instability.

One can compare these results to the other specification where interactions between cohabitation and education are included. The coefficient on cohabitation interacted with less than high school education increases in size, although it remains statistically insignificant. The positive association of premarital cohabitation with marital instability has become stronger for less educated women. On the other hand, one can see a marked decline in the coefficient of premarital cohabitation interacted with the other educational dummies. That means that the positive association of premarital cohabitation with marital instability has weakened over time only for women with a high school degree and more than a high school degree. Thus, the overall change in the coefficient on premarital cohabitation can be explained by a strong decrease in the association between cohabitation and marital instability within educational groups. Since women with more than a high school education are now the majority in the population, a change within this group mainly drives the decline in the overall coefficient on cohabitation.

I hypothesize that this finding on differences by education may be interpreted in the following: For less educated women cohabitation has always been more common than for other socioeconomic groups. For this reason, self-selection has not been as severe within this educational group even in the first cycles of the NSFG where there was a strong overall association between premarital cohabitation and marital instability. On the other hand, premarital cohabitation was relatively uncommon for well-educated women in earlier years, suggesting that the small minority of well-educated cohabiting women was perhaps more selective of divorce-prone individuals. Earlier, I discussed reasons for why marriage has become less attractive for more educated women. As more educated women have begun cohabiting, the positive self-selection effect might have mitigated. It is true that well-educated women are still less likely to cohabit in comparison to less educated women, but it is no longer uncommon. In the latest cycle, almost half of more educated women reported to have cohabited in the past or are currently cohabiting. My explanation for these findings is that cohabitation has ceased to be selective of high-risk individuals within the group of highly educated women.

Other results are in line with previous studies: Premarital conception increases the risk of marital dissolution while a marital birth decreases this risk. Religious affiliations decrease the risk of marital dissolution. Race plays a role: White and other non-Black respondents have more stable marriages than Black respondents. Respondents with an intact family background are less likely to get separated. A higher age for wives at wedding reduces the risk of separation greatly as it is found in many other empirical studies.

2.4 Causal Modeling

The proportional hazard models in the earlier section do not allow conclusions about the causal effect of premarital cohabitation on marital stability since the coefficient on premarital cohabitation is likely to be tainted by self-selection. Instrumental variables are, in general, one way to deal with endogenous regressors. Instrumental variables must be correlated with the endogenous regressor and must not be correlated with the error term in the main regression to be valid. For this reason, it is very difficult to think of a good instrument in the context of cohabitation. Since cohabitation and marriage are similar interdependent decision problems, one would not expect to find a variable satisfying the necessary condition, for in the BLS model, all variables that affect the probability of marital dissolution are also likely to affect the probability of premarital cohabitation. To the knowledge of the author, no study has attempted to implement an instrumental variable estimator in the context of cohabitation and marriage.

Other methods of causal modeling include matching on observables and random and fixed effects methods. I will use all of these methods. They are, however, no panacea in this particular case. Matching estimators only condition on observable characteristics and assume unconfoundedness. In this context, this would mean that women decide to cohabit or not before marriage independent of the effect of this decision on their marital outcomes once one conditions on the observables. This is a very strong assumption

Random and fixed effect estimators, on the other hand, rely on the presence of multiple outcomes, using data on multiple marriages and essentially differencing across marriages, correlating differences in marital dissolution with differences in premarital cohabitation. However, one may think that there are different dynamics at play in higher order marriages compared to first marriages and that these are correlated with the decision to cohabit. For example, someone might cohabit instead of marry after a divorce in order to obtain alimonies. In that case, the incentives to enter cohabitation and marriage are rather different for first and higher order marriages.

2.4.1 Matching on Observables

Matching provides a possibility of estimating the effects of premarital cohabitation on marital stability while controlling for the effects of observable variables. The independent variable is an indicator for separation for each of the first six years of the first marriage. This allows to estimate the time pattern of the effect of premarital cohabitation on marital stability. Gerfin and Lechner (2002) used a similar methodology for studying the effect of active labor market policies on labor market outcomes in Switzerland. If premarital cohabitation affects marital stability by helping the couples learn about their match quality, one would intuitively expect that the effect is strongest in the early phases of the marriage. The information advantage that could have been built up during a cohabitation vis-a-vis a couple who married right away should dissipate as the non-cohabitors learn about match quality in their marriage.

Using a matching estimator also has a second advantage. It allows for a different effect of cohabitation conditional on observable variables for each person. If there are differences in the effects of cohabitation on marital outcomes depending on the education level, then matching provides a method of dealing with these interactions.

Nearest Neighbor Matching

Nearest neighbor matching is discussed in Abadie and Imbens (2006), and the software implementation of the algorithm is discussed in Abadie, Drukker, Herr, and Imbens (2004). Premarital cohabitation is understood as a binary treatment. The treatment status is indicated by $W_i \in \{0,1\}$. For individual i one observes whether she cohabited and the outcome in year t.

$$Y_i^t = \begin{cases} Y_i^t(0) & \text{if } W_i = 0 \\ Y_i^t(1) & \text{if } W_i = 1 \end{cases} \tag{2.17}$$

where Y_i^t is a binary indicator for separation in year t of the marriage. One also observes a vector of covariates denoted by X_i. Two effects are of interest here. The average treatment effect (ATE)[4] is defined as

$$ATE = E\left(Y_i\left(1\right) - Y_i\left(0\right)\right) \tag{2.18}$$

The ATE is the average effect of the treatment on the whole population. If everyone started their co-residential union as a cohabitation, the ATE would be the effect of cohabitation on marital instability. The average treatment effect on the treated (ATT) is defined as:

$$ATT = E\left(Y_i\left(1\right) - Y_i\left(0\right)|W_i = 1\right) \tag{2.19}$$

The ATE for the subpopulation with $X = x$ can be expressed as

$$ATE = E\left[E\left(Y_i\left(1\right)|W = 1, X = x\right) - E\left(Y_i\left(0\right)|W = 0, X = x\right)\right] \tag{2.20}$$

The ATT can be expressed as

$$ATT = E\left[E\left(Y_i\left(1\right)|W = 1, X = x\right) - E\left(Y_i\left(0\right)|W = 0, X = x\right)|W = 1\right] \tag{2.21}$$

Implicitly, the treatment effects condition on survival up to t for $t > 1$. The reason for this is that one can only observe the outcome (separation or no separation) in t if the couple has not separated in any period before t. Only in the first year, there is no implicit conditioning. The more fundamental problem is that only one of the outcomes can be observed in any given year. If the individual cohabited before marriage, one cannot observe the counterfactual outcome that would have happened if the couple started their co-residential union as a marriage right away. The counterfactual outcomes therefore have to be estimated from the data.

[4]For ease of exposition, I suppress the superscripts.

The estimators are defined as follows. They are a specialization for the case of a single match.

$$\widehat{Y_i^t}(0) = \begin{cases} Y_i^t \text{ if } W_i = 0 \\ Y_j^t \text{ if } W_i = 1 \end{cases} \tag{2.22}$$

$$\widehat{Y_i^t}(1) = \begin{cases} Y_j^t \text{ if } W_i = 0 \\ Y_i^t \text{ if } W_i = 1 \end{cases} \tag{2.23}$$

for some other agent j. The counterfactual outcome is the outcome of another agent j whose observable characteristics are closest to agent i. The nearest neighbor is the neighbor whose characteristics are closest in terms of the continuous variables. The continuous variables were weighted by the inverse of their sample standard errors. An exact match for the discrete variables was used. For some agents no exact match was possible but for most this raised no issues. The ATE for period t can then be calculated as

$$ATE^t = \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{Y_i^t}(1) - \widehat{Y_i^t}(0)\right) \tag{2.24}$$

where the summation is taken over all individuals. The ATT for period t can be calculated by only summing over the treated

$$ATT^t = \frac{1}{N_{treated}}\sum_{W_i=1}\left(\widehat{Y_i^t}(1) - \widehat{Y_i^t}(0)\right) \tag{2.25}$$

The large sample properties including asymptotic standard errors are discussed in Abadie and Imbens (2006).

For all the three cycles of the NSFG the same matching variables were used, and include age at interview, age difference, race, educational achievement, religious affiliation, and fertility indicators. This is a disadvantage because there are more potentially useful variables available in the NSFG 2002 which were unavailable in earlier cycles. Gerfin and Lechner (2002) argue that for matching estimators one needs a rich data set of matching variables to satisfy the conditional independence assumption. In light of this assumption, one would not want to restrict the set of matching variables too much. However, the results seem not to differ much qualitatively depending on which set of matching variables is used.

Table 2.6 presents the ATE, ATT, and ATU for the NSFG 1988. Most of the estimates are positive and the only significant estimates are positive. These estimates are in line with earlier work and my own results confirming that there is a strong connection between cohabitation and the risk of subsequent marital instability at least for earlier years. In 1988 the proportion of surviving marriages for cohabitors was 1.2% lower after the first year than for non-cohabitors. Since about 5-6% of marriages fail within the first year, this means that cohabitors have 20% higher chances of a divorce in their first year.

Table 2.7 shows the corresponding estimates for the NSFG 1995. Even though most of the estimates are still positive according to these estimates, at least the ATT has changed. In the second year there is now a marked reduction in risk of separation for cohabitors. Also, it seems that if there is a positive relationship between cohabitation and the risk of marital instability, this relationship weakens as compared to the corresponding estimates for the previous cycle.

Table 2.8 shows the ATE, ATT, and ATU for the year 2002. Comparing to the estimates of the year 1988, one can clearly see that now premarital cohabitation decreases marital instability for the group of cohabitors. For the whole population, cohabitation does not have any effect in either direction.

My main question here was whether the effect of premarital cohabitation on marital stability has changed over time. To answer this question one can calculate the difference in the treatment effects between 2002 and 1988 and see whether they are significantly different from zero using t-tests since these two samples are independently drawn. As shown in table 2.9, most differences in the treatment effects are negative meaning that the positive effect of cohabitation on marital instability has weakened or even turned its sign. Furthermore, only negative differences appear to be statistically significant. I conducted similar tests calculating the difference in the treatment effects 2002-1995 and 1995-1988. A similar pattern emerges here. The apparent positive relationship between premarital cohabitation and marital instability has weakened.

2.4.2 Random Effects Model

The Lillard, Brien, and Waite (1995) model[5]

LBW model the decision to cohabit before marriage and the marital dissolution process simultaneously. There is an unobserved heterogeneity term in both of these processes that may be correlated. This heterogeneity term is assumed to be perma-

[5] LBW henceforth.

nent for a person so that the correlation can be identified by using multiple marriage outcomes for a person. A positive correlation between the heterogeneity terms indicates self-selection of individuals with a high risk of marital disruption into premarital cohabitation. In this model, the coefficient on premarital cohabitation can be interpreted as the causal effect of cohabitation on marital stability since it is purged from any self-selection if the model is correct. I discuss the marital dissolution process and the decision to cohabit separately.

Marital Dissolution LBW use a continuous time duration model of marriage. The logarithm of the instantaneous probability of dissolution at time t for the m^{th} marriage conditional on not having dissolved before that time (log-hazard) is given by the following equation:

$$\begin{aligned} \ln h_m\left(X_m^d, Coh_m, t, \delta\right) &= \alpha_0 + \alpha_1^{'} DurMar\left(t\right) + \alpha_2^{'} DurBirth\left(t\right) + \\ &\quad \alpha_3^{'} X_m^d + \alpha_4^{'} Coh_m + \delta \end{aligned} \tag{2.26}$$

In this equation $DurMar$ and $DurBirth$ represent duration splines starting at the beginning of the marriage and the birth of the first child respectively. In contrast to the proportional hazard model discussed earlier, this model uses a completely parametric form for the duration dependency. The time-dependent part can also be written as:

$$h_{0m}\left(t\right) = \exp\left(\alpha_0 + \alpha_1^{'} DurMar\left(t\right) + \alpha_2^{'} DurBirth\left(t\right)\right) \tag{2.27}$$

This is the baseline hazard. All the other regressors will shift this baseline hazard proportionally. The baseline survivor function is given as:

$$S_{0m}(t) = \exp\left(-\int_{t_0}^{t} h_{0m}(t)\, dt\right) \tag{2.28}$$

The survival function is then given as:

$$S_m\left(X_m^d, Coh_m, t, \delta\right) = \prod_{i=1}^{I} \left(\frac{S_{0m}(t_{i+1})}{S_{0m}(t_i)}\right)^{-\exp\left(\alpha_3' X_m^d + \alpha_4' Coh_m + \delta\right)} \tag{2.29}$$

where I is the number of periods in which the covariates are constant. Since only a dummy for a marital birth is time-varying the maximum number of these periods is two. The regressor set X_m^d includes regressors that are fixed for a given marriage but may vary across one individual's marriages including dummies for higher order marriages, education, age at wedding, and other socioeconomic variables. The coefficient α_4 measures the effect of premarital cohabitation (with the future spouse). Finally, there is an unobserved component δ which is assumed to be fixed for all marriages of a given woman.

It would be possible to estimate this model using maximum likelihood:

$$-\int_{\delta} \frac{1}{\sigma_\delta}\phi\left(\frac{\delta}{\sigma_\delta}\right) \prod_{m=1}^{M} S_m\left(X_m^d, Coh_m, t, \delta\right) h_m\left(X_m^d, Coh_m, t, \delta\right)^{D_m} d\delta \tag{2.30}$$

This is a random effects model where some personal characteristic of the women is the same across all marriages. Implicitly, it is assumed that censoring is exogenous or random. However, in our case these assumptions are likely to be violated.

Cohabitation The decision to cohabit before the m^{th} marriage is modeled as a probit model:

$$I_m = \beta_0 + \beta_1^{'} X_m^c + \epsilon + \eta_m \tag{2.31}$$

$$Coh_m = \begin{cases} 1 \text{ if } I_m > 0 \\ 0 \text{ otherwise} \end{cases} \tag{2.32}$$

The set of regressors X_m^c again includes socioeconomic variables, ϵ is again the unobserved heterogeneity. It is constant across all marriages of an individual. Finally, η_m is distributed i.i.d. according to a standard normal distribution.

The Joint Process The heterogeneity components δ and ϵ are assumed to be drawn from a bivariate normal distribution. That is

$$\begin{pmatrix} \delta \\ \epsilon \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\delta^2 & \sigma_{\delta\epsilon} \\ \sigma_{\delta\epsilon} & \sigma_\epsilon^2 \end{pmatrix} \right) \tag{2.33}$$

A positive correlation, that is $\sigma_{\delta_\epsilon} > 0$, would indicate self-selection of individuals with a high risk of marital disruption into premarital cohabitation The unobserved heterogeneity is integrated out so that the joint marginal likelihood contribution of all marriages of a given woman is then given by:

$$\int_\delta \int_\epsilon \frac{1}{\sigma_\delta \sigma_\epsilon} \phi \left(\frac{\delta}{\sigma_\delta}, \frac{\epsilon}{\sigma_\delta} \mid \rho_{\delta\epsilon} \right) \prod_{m=1}^{M} \Big[S_m \left(X_m^d, Coh_m, t, \delta \right) h_m \left(X_m^d, Coh_m, t, \delta \right)^{D_m} \tag{2.34}$$
$$\times \Phi \left((2Coh_m - 1) \left(\beta_0 + \beta_1^{'} X_m^c + \epsilon \right) \right) \Big] \, d\delta d\epsilon$$

where the product is taken over the first three marriages of a woman. $D^m = 1$ indicates a completed marriage spell while $D^m = 0$ indicates a censored spell. Conditional on the unobserved heterogeneity and the covariates there is no correlation in outcomes across marriages for a given woman. The unobserved heterogeneity and the correlation between the heterogeneity component is identified even without exclusion restrictions since one observes more than one marriage for some women. Notice that this identification strategy relies heavily on the assumed functional form, namely that the unobserved heterogeneity components are drawn from a bivariate normal distribution. After controlling for other exogenous covariates and this heterogeneity, the coefficient on cohabitation should reveal the true structural effect of cohabitation on marital instability. If it is found that this coefficient is not different from zero, the finding would lend support to the hypothesis that the observed association of premarital cohabitation with marital instability is due to self-selection. The correlation between the heterogeneity components sheds light on whether the process of self-selection into cohabitation and marriage may have changed between the cohorts of women studied in LBW and the cohort of women investigated in the NSFG 2002. LBW found a strong and positive correlation between these heterogeneity components using the National Longitudinal Study of the High School Class of 1972 with its follow-up in 1986. I estimate the LBW model using the NSFG 2002 dataset with a similar set of covariates as LBW and interpret differences in findings as due to time trends in the selection effect. Unfortunately, detailed cohabitation data relative to all higher order marriages was not collected for the NSFG 1988 and 1995. Therefore, I cannot study the time-evolution of the correlation between the heterogeneity components with my datasets.

There are several problems with this model. I already mentioned that the identification strategy heavily relies on functional forms not on exclusion restrictions. Although LBW exclude some variables from each of the processes, one could argue that the theoretical justification for doing so is not always convincing. For example, they exclude the sex ratio from the marital dissolution process. However, the sex ratio might determine the outside option outside of the present marriage (after divorce and potential remarriage). Similarly, they exclude the age difference between the spouses from their cohabitation process. One might argue that a higher age difference between spouses leads to cohabitation because the incentive to gather information might be bigger if the age background of the spouses is different. Also, they assume that censoring is exogenous, leading to two problems. First, some of their covariates are likely to influence the probability that a spell is censored (covariate-dependent censoring). Furthermore, the spells are not censored independently (random independent censoring). If the first spell is very long, one is more likely to see subsequent spells censored.

Comparison to my Results

LBW explicitly modeled the decision to start a co-residential union as a cohabitation and the hazard of marital dissolution jointly. They found a strong statistically significant relationship between the unobserved heterogeneity in both processes. More divorce-prone individuals are more likely to cohabit. The causal effect of cohabitation on marital dissolution was statistically insignificant. I was unable to include all of their variables due to data restrictions but I tried to be as close as possible to their estimates. These results are shown in table 2.10 for the marital dissolution process and table 2.11 for the probit models.

In the first two columns of table 2.10, the results for a parametric hazard model without heterogeneity are shown. Coefficients are again reported as hazard ratios where a coefficient greater than one indicates that the variable increases marital instability while a coefficient smaller than one indicates a reduction in the hazard of marital dissolution. In this specification, the coefficient on premarital cohabitation is below one, indicating that premarital cohabitation decreases marital instability. I included estimates with the previous cohabitation (with someone other than the husband) dummy and without that dummy since LBW did not have that dummy. One could expect that the inclusion of this variable would change the coefficient on premarital cohabitation. However, this turns out to be not the case. Duration dependency was introduced by using splines starting at the date of the wedding and after the birth of the first child during marriage. All the other variables included are the same variables as in the proportional hazard regressions from earlier sections. One difference is that the standard errors are smaller than for the proportional hazard regressions since a fully parametric model is estimated in contrast to the semiparametric proportional hazard model.

In columns III and IV the model with unobserved heterogeneity is estimated but without simultaneously modeling the probit equation for premarital cohabitation. Including unobserved heterogeneity reduces the effect that a higher order marriage has on marital instability. LBW partly explain why higher order marriages are less stable: Due to the selection of individuals with higher risk of marital dissolution into these marriages, their unions are more likely to get dissolved. In contrast to their results, including an unobserved heterogeneity component does not affect the significance or the quantitative importance of premarital cohabitation. However, it changes the estimated duration dependency of the marital dissolution process.

Finally, in columns V and VI the estimates for the marital dissolution process are shown when the probit process is modeled simultaneously. Due to numerical difficulties, I restricted the standard deviation of the heterogeneity component in the marital dissolution process and in the probit equation to be equal. LBW estimated these standard deviations to be roughly equal. The coefficient on premarital cohabitation stays below one even in this specification meaning that premarital cohabitation decreases marital instability. The estimated correlation between the unobserved heterogeneity components is negative although very imprecisely estimated. This would imply a self-selection of women into cohabitation who are more likely to have stable marriages, ceteris paribus.

This contrasts sharply with the results in LBW who found that the correlation between these heterogeneity components has a positive sign. Furthermore, they found that when estimating the probit and the dissolution process simultaneously the coefficient on premarital cohabitation loses its significance. Their results imply that there is no true causal effect of cohabitation on marital instability. My results, in contrast, imply that there is a causal effect of cohabitation. It increases the stability of marriages. There is no self-selection of divorce-prone individuals into cohabitation.

I interpret these findings as indicative of a change in the self-selection process into premarital cohabitation. The time trend of increasing cohabitation rates has also changed the process of self-selection into premarital cohabitation.

Table 2.11 presents the estimation results for the probit process. Column I is a probit model without heterogeneity, column II is a model with heterogeneity, and column III shows the parameter estimate for the probit process when it is simultaneously estimated with the marital dissolution process. Column III presents also the coefficients of the probit process for the simultaneous estimation and corresponds to column V of table 2.10. The results confirm many previous results about the factors leading to premarital cohabitation.

2.4.3 Fixed Effects Models

The mixed proportional hazard model is an extension of the proportional hazard model introducing unobserved heterogeneity. It was developed by Lancaster (1979) and a recent survey can be found in Van den Berg (2001). This model will serve as a building block for the following discussion of fixed effects model in the context of multiple duration models.

The hazard in the mixed proportional hazard model (Van den Berg 2001) is defined as:

$$\theta\left(t|x,v\right)=\psi\left(t\right)\theta_0\left(x\right)v \tag{2.35}$$

This can be specialized to:

$$\theta\left(t|x,v\right)=\psi\left(t\right)\exp\left(x'\beta\right)v \tag{2.36}$$

This is an extension of the proportional hazard model. The additional component, v, is an unobserved heterogeneity component. One way of dealing with this unobserved heterogeneity would be to make additional distributional assumptions on it similar to the ones by LBW. Another way would be to treat them as incidental parameters. This approach is discussed here. Van Den Berg (2001) shows that the mixed proportional hazard model can also be written as:

$$\log\int_0^t \psi\left(u\right)du=-x'\beta-\log v+\varepsilon \tag{2.37}$$

where the ε are extreme value type 1 distributed. Sometimes, it is possible to observe two spells for one person.

$$\log\int_0^{t_1} \psi\left(u\right)du \;=\; -x_1'\beta-\log v+\varepsilon_1 \tag{2.38}$$

$$\log\int_0^{t_2} \psi\left(u\right)du \;=\; -x_2'\beta-\log v+\varepsilon_2 \tag{2.39}$$

where the $\varepsilon_1,\varepsilon_2$ are iid EV1 distributed. This representation of the model suggests treating the $\log v$ as fixed effects. Since the hazard $\psi\left(u\right)$ has to be non-negative, the left hand side of this equation is a positive monotone transformation of the duration variable. Recently, there has been some research on the use of rank estimation models in the context of a panel model (see for example, Abrevaya 1999, 2000; Khan and Tamer 2006). These estimation methods are using the fact that left hand side of equation 2.36 is a positive monotone transformation without specifying them exactly. A problem with these estimators is that their computation is rather difficult since the objective function is discontinuous. Smoothing provides a partial solution to this problem. Censoring poses a serious problem for these estimators since for many women, one does not even observe the start of their second marriage and for many women the second spell is censored. More problematically, the fixed effect introduces correlation in the expected durations of each of the spells and therefore the usual assumption of independent censoring is likely to be violated.

I have tested rank estimators of this type but found that in my sample the procedure produced large standard errors leaving all estimates very unreliable. For this reason, I do not report those results here. In all of these models, the standard errors on the coefficient on premarital cohabitation were so large that one cannot reject the null of no effect of cohabitation.

Lee (2003) proposes an estimator for estimating a panel fixed effect duration model with an unknown transformation of the linear index and dependent right censoring. Dependent right censoring is particularly important in the case of the NSFG 2002 dataset. Intuitively, one is more likely to observe higher order marriages of women whose first marriage was unstable. Similar to the rank estimations of the transformation model, Lee (2003) proposes an estimation technique for the mixed proportional hazard model with person-specific effects. He takes the following transformation model as starting point:

$$H_i\left(dur_{it}\right) = X_{it}'\beta + U_i + \varepsilon_{it} \tag{2.40}$$

where $H_i\left(dur_{it}\right)$ is a person-specific unknown monotonic transformation of the variable, U_i is the fixed-effect, and ε_{it} is the i.i.d. error term. I consider the panel case since I have observations on first and second marriages. Lee (2003) discusses in detail a censoring mechanism that needs to be adopted to my case. He assumes that after the first duration ends, the second duration immediately follows. However, after a separation there is typically some waiting period until a person marries again. The censoring mechanism in my case could be described as follows: There are durations dur_1 and dur_2 as well as a random period between the marriages w. The censoring time C is random. This assumption is violated since the date of the interview is fixed. However, to my knowledge there are no fixed effect estimators of this model dealing with the problem of fixed censoring time.

The following cases are possible:

1. $C \geq dur_1 + w + dur_2$: both durations are observed without censoring.

2. $dur_1 \leq dur_1 + w \leq C \leq dur_1 + w + dur_2$: the first duration is observed without censoring. The second duration is observed censored.

3. $dur_1 \leq C \leq dur_1 + w$: the first duration is observed without censoring. The second duration is not observed. This is the case when one observes a woman who separated but never married again.

4. $C \leq dur_1$: only the first duration is observed censored.

Similar to the arguments in Lee (2003), one can see why the assumption of independent random censoring is violated. Notice that the first duration is censored by $C \equiv C_1$. The second duration is censored by $C_2 \equiv (C_1 - dur_1 - w)\, 1\,(dur_1 \leq C_1)$. Both durations are correlated because of the fixed effect. But then C_2 is also correlated with dur_2. An observer is more likely to observe both durations uncensored for women whose fixed effect causes them to have less stable marriages. Instead of observing the duration, the econometrician observes a pair (Y_j, Δ_j) where $Y_j = \min(dur_j, C_j)$ for $j = 1, 2$. The observed data then consists of i.i.d. realizations of $\{(Y_{i1}, Y_{i2}, X_{i1}, X_{i2}, \Delta_{i1}, \Delta_{i2}) : i = 1, 2, ..., N\}$. Let $G(c)$ denote the survivor function of C, that is $\Pr(C \geq c)$. Let $\Delta X_i = X_{i1} - X_{i2}$. I assume that the ε_{it} are i.i.d. draws from an EV1 distribution. Let $L(u) = \int\limits_{-\infty}^{\infty} (1 - F(u + v))\, dF(v)$, where $F(v)$ is the cdf of the extreme value distribution. Let $l(u) = \frac{-dL(u)}{du}$. Lee (2003) shows that under his assumptions the following moment condition holds:

$$E\left(w_h(\Delta X'\beta)\, \Delta X \frac{\Delta_1 \Delta_2}{G(Y_1 + w + Y_2)} \left[1(Y_1 > Y_2) - L(\Delta X'\beta)\right]\right) = 0 \tag{2.41}$$

where w_h is some weight function. This suggests the following estimating equation:

$$N^{-1} \sum_{i=1}^{N} \left(w_h\left(\Delta X_i' \widehat{\beta}\right) \Delta X \frac{\Delta_1 \Delta_2}{G_N(Y_1 + w + Y_2)} \left[1(Y_1 > Y_2) - L\left(\Delta X_i' \widehat{\beta}\right)\right]\right) = 0 \tag{2.42}$$

where $G_N(.)$ is the Kaplan-Meier estimator of the survival function. The expression $\frac{\Delta_1 \Delta_2}{G_N(Y_1 + w + Y_2)}$ assigns weight to individuals with two uncensored spells proportional to the inverse of the probability of observing those spells uncensored. All the other observations enter only in the estimate of the survival function. This is particularly important in my case. If a woman does not start a second marriage, many covariates like the age at the second marriage will not be observed either. One can further specialize this estimating equation by assuming the following weight function $w_h = \frac{l(u)}{L(u)(1-L(u))}$. Lee (2003) shows that this is equivalent to the weighted logit estimator of the form:

$$\max_{\widehat{\beta}} N^{-1} \sum_{i=1}^{N} W\left(\left[1\left(Y_1 > Y_2\right) \log\left(L\left(\Delta X'\beta\right)\right) + 1\left(Y_1 \leq Y_2\right) \log\left(1 - L\left(\Delta X'\beta\right)\right)\right]\right) \tag{2.43}$$

where $W = \frac{\Delta_1 \Delta_2}{G_N(Y_1 + w + Y_2)}$. In the tables, I will present the results for the weighted logit estimator but will use the more standard notation for $\Delta X_i = X_{i2} - X_{i1}$. Lee (2003) shows that this estimator is $N^{-\frac{1}{2}}$-consistent and asymptotically normal.

For this estimator [6], in a first step the survival function for c, the censoring time, is estimated using a Kaplan-Meier estimator. Table 2.12 shows the results. The independent variable is a dummy which is 1 if the second spell is longer than the first one. In the dataset there are 228 uncensored (completed) first and second spells on which the estimates are based. Notice, however, that the survivor function for the censoring variable c was estimated using the full sample.

In column I the squared age of the wife at the wedding is included. The coefficient on premarital cohabitation is greater than zero indicating a *positive* effect of

---

[6] I am grateful to Sokbae Lee for sending me his Gauss code of this estimator.

cohabitation on marital *stability* or a *negative* effect on marital *instability*. One can see that there seems to be a non-linearity in the effect of the wife's age. The wife's age at the wedding is positively related to marital stability while the squared age is negatively related to marital stability. Earlier works often include only the linear term when investigating first marriages where the non-linearity might not be an issue. The husband's age at wedding is positively related to marital stability, though insignificantly. Premarital cohabitation is positively related to marital stability in all specifications at least on the 10% significance level. Surprisingly, the trend variable seems to be positive. That would mean that second marriages are more stable than first marriages if one controls for a person-level fixed effect. LBW find that when they control parametrically for person-level heterogeneity (model 2 in table 4 of their paper) that higher-order marriages are more stable than first marriages. However, this particular coefficient is not significant on the 10% level, in contrast to my result.

2.5 Conclusion

In this essay, I have reassessed the question of the influence of premarital cohabitation on marital instability. A theoretical search model of marriage and cohabitation suggests that cohabitation should help couples learn about their match quality and should decrease their dissolution rates. On the other hand, this model also shows that the average match quality of couples who transform their cohabitation into a marriage is lower than for couples who marry right away. This self-selection of high-risk individuals could explain the empirical evidence that has been established for the US and other industrialized countries showing that marriages preceded by cohabitation are less stable. I also show how the process of self-selection changes when the underlying benefits to marriage and cohabitation change.

This essay demonstrates that the once-strong association between premarital cohabitation and marital instability has weakened over the three recent cycles of the NSFG, and there is no longer an association in the most recent data. A strong decline in the correlation of cohabitation with marital dissolution within the group of more educated women drives the result. One does not observe a similar change in the coefficient on cohabitation within the group of less educated women, and the coefficient of premarital cohabitation on marital instability was never strong within the group of less educated women. My suggested explanation for this result is that there was no strong self-selection effect on unobservables within the group of less educated women because it was always more common for this group to cohabit. For more educated women, however, it was less common to cohabit before marriage in earlier years. Cohabiting women might have been a more select group in the past, and the argument that they have elevated risk of marital dissolution because of unobservable factors might have applied to them. In more recent years, the incentives to cohabit and marry have changed for this group and a higher proportion has begun cohabiting. While it is still less common for highly educated women to cohabit, it is no longer a fringe phenomenon. For this reason, I hypothesize that the self-selection within the group of highly educated women has declined. This, in turn, affects the measured association between premarital cohabitation and marital instability. Using the LBW model, I no longer find significant self-selection strengthening my interpretation.

While my results are new and surprising for the United States, they are in line with more recent evidence from Denmark (Svarer 2004) and other European countries (Liefbroer and Dourleijn 2006). When about half of the population cohabits, cohabitation ceases to be selective of divorce-prone individuals. Another reason that possibly puts doubt on the thesis of a stable relationship between premarital cohabitation and marital instability is that the character of cohabitation might have changed over time. This change in character of cohabitation might be badly measured by a binary indicator of cohabitation. Cherlin (2004) cites work by the British demographer Kiernan suggesting that acceptance of cohabitation in society follows four steps. In the first step, cohabitation is a fringe phenomenon, after which it becomes acceptable as testing ground for marriage in a second step. In step three, cohabitation becomes an accepted alternative to marriage, and finally it even becomes virtually indistinguishable from marriage. If this thesis is true, then it would be even more likely that the self-selection process into cohabitation has changed fundamentally. Given the rapid changes in marriage and cohabitation behavior in the United States, the relationship between cohabitation and marital instability may not be stable yet.

Table 2.1: Trends by Age in the Percentage Ever Cohabiting and Currently Cohabiting of Not Currently Married

	Percentage ever cohabiting			Percentage currently cohabiting		
Age	1987	1995	2002	1987	1995	2002
19-24	29	38	38	14	15	19
25-29	41	47	66	20	21	26
30-34	40	49	68	17	21	21
35-39	30	48	72	11	17	19
40-44	22	41	66	14	13	20
Total	33	45	56	15	17	21

Notes: Percentages for years 1987 and 1995 are taken from table 1 in Bumpass and Lu (2000). Percentages for 2002 based on own calculations using NSFG 2002.

Table 2.2: Percentage of Women Aged 19-44 Who Have Ever Cohabited

	Percentage of women who have ever cohabited		
	1987	1995	2002
Education			
Less than high school	43	59	68
High school	32	46	66
Some College	30	39	51
College 4+	31	37	46
Race/Ethnicity			
White non-Hispanic	32	45	57
Black	36	45	60
Hispanic	30	39	54

Notes: Percentages for years 1987 and 1995 are taken from table 2 in Bumpass and Lu (2000). Percentages for 2002 based on own calculations using NSFG 2002.

Table 2.3: Means of Variables Relative to First Marriage

	Non-cohabitors			Cohabitors		
	1988	1995	2002	1988	1995	2002
Percentage in population	72.8	64.1	54.3	27.2	35.9	45.7
1st marriage intact[a] at time of interview (%)	67.5	65.5	65.0	67.5	66.1	66.2
Less than high school (%)	14.7	15.5	15.0	17.6	19.3	19.7
High school (%)	39.4	36.8	25.1	35.7	34.8	24.9
More than high school (%)	45.9	47.7	59.9	46.7	45.9	55.5
Age at interview	33.6	33.7	35.4	31.4	32.5	33.7
Age at wedding (wife)	21.1	22.4	22.4	22.4	21.0	24.3
Age at wedding (husband)	23.8	24.3	24.5	26.0	24.2	26.9
Age difference	2.7	2.2	2.2	3.6	3.2	2.6
Premarital conception (%)	27.4	31.9	31.9	43.2	53.2	56.5
Premarital lifebirth (%)	7.1	10.0	10.0	14.7	23.6	29.3
Marital lifebirth (%)	75.3	72.3	72.3	58.4	60.7	60.6

Notes: Sample weights are used.

a: First marriage is still intact at the time of the interview or first marriage was ended by the death of the spouse.

Table 2.4: Age, Marital, and Cohabitation Status by Education Group

	Less than high school			High school			More than high school		
	1988	1995	2002	1988	1995	2002	1988	1995	2002
Cohabited (%)	30.9	41.0	52.5	25.3	34.6	45.5	27.5	35.0	43.9
Marriage intact[a]	56.9	53.3	57.8	66.2	63.9	56.8	72.1	71.6	71.6
Of cohabitors:									
Marriage intact[a]	65.0	56.8	55.3	60.9	66.4	59.5	73.5	69.9	73.0

Notes: Sample weights are used.

a: First marriage is still intact at the time of the interview or first marriage was ended by the death of the spouse.

Table 2.5: Proportional Hazard Regressions. Dep. Var. Hazard of Separation of First Marriage

Variable	1988	1995	2002	1988	1995	2002
Premarital	1.486***	1.334***	1.123			
cohabitation	(0.109)	(0.070)	(0.096)			
No HS*				1.123	1.153	1.320
prem. cohabit.				(0.180)	(0.109)	(0.229)
High School*				1.900***	1.348***	1.090
prem.cohabit.				(0.194)	(0.109)	(0.159)
More than HS*				1.341***	1.437***	1.069
prem. cohabit.				(0.141)	(0.109)	(0.116)
Education						
No HS	0.969	1.084	0.918	1.027	1.178*	0.828
	(0.091)	(0.075)	(0.102)	(0.110)	(0.099)	(0.128)
High School	0.937	0.991	1.239**	0.855**	1.015	1.229
	(0.063)	(0.053)	(0.120)	(0.066)	(0.065)	(0.163)
Fertility Var.						
Premarital	1.530***	1.378***	1.711***	1.540***	1.376***	1.721***
conception	(0.109)	(0.070)	(0.167)	(0.110)	(0.077)	(0.167)
Premarital	1.001	1.245***	1.031	1.008	1.260***	1.022
lifebirth	(0.110)	(0.096)	(0.119)	(0.111)	(0.098)	(0.118)
Marital	0.580***	0.778***	0.726***	0.575***	0.776***	0.729***
lifebirth	(0.044)	(0.047)	(0.065)	(0.044)	(0.047)	(0.065)
Religion						
Protestant	0.694***	0.760***	0.824	0.682***	0.763***	0.821
	(0.089)	(0.056)	(0.104)	(0.087)	(0.056)	(0.104)
Catholic	0.655***	0.698***	0.823*	0.648***	0.698***	0.820*
	(0.089)	(0.056)	(0.088)	(0.088)	(0.056)	(0.088)
Jewish	0.439***	0.794		0.424***	0.788	
	(0.143)	(0.184)		(0.139)	(0.183)	
Other	0.961	0.629**	0.934	0.961	0.626**	0.932
	(0.231)	(0.138)	(0.167)	(0.230)	(0.138)	(0.167)
No spec. den.		0.781			0.776	
		(0.122)			(0.122)	

Table cont'd on following page

Table 2.5 cont'd

Variable	1988	1995	2002	1988	1995	2002
Race						
White	0.798***	0.754***	0.822**	0.807***	0.759***	0.818*
	(0.060)	(0.052)	(0.086)	(0.061)	(0.052)	(0.085)
Other	0.523***	0.652***	0.646**	0.525***	0.657***	0.640***
	(0.106)	(0.082)	(0.112)	(0.105)	(0.082)	(0.111)
Family Background						
Intact fam.	0.815***			0.818***		
	(0.062)			(0.063)		
Adopted child		1.715**			1.708**	
		(0.414)			(0.415)	
No int. family		1.251***	1.384***		1.254***	1.378***
		(0.062)	(0.126)		(0.062)	(0.125)
Wife's age at wedding	0.879***	0.913***	0.906***	0.880***	0.913***	0.905***
	(0.012)	(0.009)	(0.014)	(0.012)	(0.009)	(0.013)
Husband's age at wedding	1.030***	0.994	1.005	1.030***	0.994	1.006
	(0.006)	(0.006)	(0.008)	(0.006)	(0.006)	(0.008)
Number of obs.	5030	6776	4030	5030	6776	4030

Notes: Sample weights are used. Estimates reported as hazard ratios. A coefficient of greater than one indicates an increase in the hazard of marital dissolution while a coefficient of smaller than one indicates a decrease in the hazard.

Standard errors in parentheses. *** 1% ** 5% *10% significant different from 1.

Table 2.6: Nearest Neighbor Matching - NSFG 1988. Dependent Variable: Indicator for Separation of First Marriage in Each of the First Six Years

Year	Avg. Treatment Effect	Avg. Treatment Effect (Treated)	Avg. Treatment Effect (Untreated)
1	0.006	0.002	0.008
	(0.009)	(0.011)	(0.010)
2	0.024***	0.013	0.028***
	(0.010)	(0.012)	(0.012)
3	0.006	-0.005	0.010
	(0.011)	(0.013)	(0.012)
4	0.019*	0.005	0.024**
	(0.012)	(0.014)	(0.014)
5	0.016	0.025*	0.013
	(0.012)	(0.014)	(0.013)
6	0.002	-0.007	0.004
	(0.012)	(0.015)	(0.013)

Notes: Sample weights are used. Matching variables include age at interview, age difference, intact family background, race, and religion.
Standard errors in parentheses. *** 1% ** 5% * 10% significance levels.

Table 2.7: Nearest Neighbor Matching - NSFG 1995. Dependent Variable: Indicator for Separation of First Marriage in Each of the First Six Years

Year	Avg. Treatment Effect	Avg. Treatment Effect (Treated)	Avg. Treatment Effect (Untreated)
1	-0.004	-0.006	-0.003
	(0.007)	(0.009)	(0.008)
2	-0.006	-0.020*	0.002
	(0.009)	(0.011)	(0.010)
3	0.013	0.009	0.015*
	(0.008)	(0.010)	(0.009)
4	0.015*	0.034***	0.006
	(0.008)	(0.010)	(0.009)
5	-0.001	0.006	-0.004
	(0.008)	(0.010)	(0.009)
6	0.010	0.018*	0.006
	(0.008)	(0.010)	(0.009)

Notes: Sample weights are used. Matching variables include age at interview, age difference, intact family background, race, and religion.
Standard errors in parentheses. *** 1% ** 5% * 10% significance levels.

Table 2.8: Nearest Neighbor Matching - NSFG 2002. Dependent Variable: Indicator for Separation of First Marriage in Each of the First Six Years

Year	Avg. Treatment Effect	Avg. Treatment Effect (Treated)	Avg. Treatment Effect (Untreated)
1	-0.011	-0.033***	0.009
	(0.011)	(0.013)	(0.013)
2	0.001	0.005	-0.001
	(0.010)	(0.013)	(0.012)
3	0.010	-0.008	0.024*
	(0.011)	(0.013)	(0.014)
4	0.008	0.007	0.009
	(0.013)	(0.014)	(0.016)
5	-0.004	-0.023	0.009
	(0.012)	(0.015)	(0.014)
6	-0.009	-0.040**	0.013
	(0.014)	(0.017)	(0.017)

Notes: Sample weights are used. Matching variables include age at interview, age difference, intact family background, race, and religion.
Standard errors in parentheses. *** 1% ** 5% * 10% significance levels.

Table 2.9: Change in Treatment Effects - NSFG 1988, 1995, and 2002

	Difference 1995-1988			Difference 2002-1995			Difference 2002-1988		
Year	ATE	ATT	ATU	ATE	ATT	ATU	ATE	ATT	ATU
1	-0.010	-0.008	-0.011	-0.007	-0.027*	0.011	-0.017	-0.035**	0.001
	(0.011)	(0.014)	(0.013)	(0.013)	(0.016)	(0.015)	(0.014)	(0.017)	(0.016)
2	-0.030**	-0.034**	-0.026*	0.007	0.025	-0.003	-0.023	-0.009	-0.029*
	(0.013)	(0.016)	(0.015)	(0.014)	(0.017)	(0.016)	(0.015)	(0.017)	(0.029)
3	0.007	0.013	0.005	-0.003	-0.016	0.009	0.004	-0.003	0.014
	(0.014)	(0.017)	(0.015)	(0.014)	(0.017)	(0.017)	(0.016)	(0.018)	(0.019)
4	-0.004	0.029*	-0.018	-0.007	-0.026	0.003	-0.011	0.003	-0.015
	(0.015)	(0.017)	(0.017)	(0.015)	(0.017)	(0.018)	(0.018)	(0.020)	(0.021)
5	-0.017	-0.019	-0.017	-0.004	-0.029	0.012	-0.020	-0.048**	-0.005
	(0.015)	(0.017)	(0.016)	(0.015)	(0.018)	(0.017)	(0.017)	(0.020)	(0.020)
6	0.008	0.025	0.002	-0.018	-0.058***	0.007	-0.010	-0.033	0.009
	(0.015)	(0.018)	(0.016)	(0.016)	(0.020)	(0.019)	(0.019)	(0.023)	(0.437)

Notes: Estimates of ATE, ATT, and ATU are based on tables 2.6, 2.7, and 2.8
Standard errors in parentheses. *** 1% ** 5% * 10% significance levels.

Table 2.10: Maximum Likelihood Estimation of Duration Model with and without Heterogeneity. Dependent Variable: Hazard of Separation of First Three Marriages

Variable	I	II	III	IV	V	VI
Premarital	0.856***	0.851***	0.822***	0.816***	0.819***	0.812***
cohabit.	(0.050)	(0.050)	(0.061)	(0.061)	(0.062)	(0.062)
Previous	1.276***		1.391***		1.403***	
cohabit.	(0.065)		(0.086)		(0.090)	
Time since marriage						
Constant	0.130***	0.134***	0.147***	0.147***	0.148***	0.152****
	(0.267)	(0.268)	(0.312)	(0.312)	(0.317)	(0.316)
Months						
0-12	1.000	1.000	1.012	1.012	1.013	1.013
	(0.013)	(0.013)	(0.014)	(0.014)	(0.013)	(0.013)
13-48	1.001	1.001	1.007**	1.007**	1.008**	1.008***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
49-120	0.996***	0.996***	1.000	1.000	1.000	1.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
120	0.997*	0.997*	1.000	1.000	1.000	1.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Time since marital birth						
Months						
0-12	0.107***	0.105***	0.087***	0.087***	0.085***	0.084***
	(0.763)	(0.763)	(0.766)	(0.766)	(0.769)	(0.769)
13-24	0.994	0.994	0.992	0.992	0.992	0.991
	(0.072)	(0.072)	(0.072)	(0.072)	(0.073)	(0.073)
>24	1.079***	1.079***	1.075***	1.075***	1.075***	1.075***
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Premarital	2.010***	2.049***	2.221***	2.275***	2.244***	2.304***
conception	(0.059)	0.058	(0.073)	(0.072)	(0.073)	(0.073)
Premarital	0.721***	0.722***	0.729***	0.726***	0.727***	0.724***
birth	(0.080)	(0.081)	(0.096)	(0.096)	(0.098)	(0.098)
Education						
No HS	1.055	1.058	1.060	1.065	1.062	1.068
	(0.066)	(0.065)	(0.085)	(0.084)	(0.088)	(0.087)
High school	1.152***	1.157***	1.182**	1.181**	1.187**	1.186**
	(0.055)	(0.056)	(0.072)	(0.071)	(0.073)	(0.073)

Table cont'd on following page

Table 2.10 cont'd

	I	II	III	IV	V	VI
Race						
White	1.004	1.008	1.025*	1.026*	1.027	1.026
	(0.124)	(0.124)	(0.145)	(0.144)	(0.148)	(0.148)
Black	1.244	1.240	1.358	1.346	1.368*	1.356*
	(0.139)	(0.139)	(0.167)	(0.166)	(0.171)	(0.170)
Religion						
Protestant	0.869	0.837	0.848	0.818	0.845	0.815
	(0.128)	(0.126)	(0.162)	(0.159)	(0.167)	(0.164)
Catholic	1.104	1.050	1.118	1.060	1.118	1.061
	(0.129)	(0.128)	(0.165)	(0.161)	(0.170)	(0.167)
No religion	1.048	1.023	1.033	1.016	1.030	1.016
	(0.137)	(0.136)	(0.175)	(0.172)	(0.181)	(0.178)
Family Background						
Int. family	0.715***	0.706***	0.677***	0.667***	0.672***	0.662***
	(0.050)	(0.050)	(0.066)	(0.066)	(0.067)	(0.067)
Not born	0.662***	0.661***	0.629***	0.628***	0.625***	0.622***
in US	(0.079)	(0.079)	(0.099)	(0.099)	(0.100)	(0.100)
Wife						
Age at wedd.	0.895***	0.897***	0.879***	0.881***	0.877***	0.880***
	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)
Has kids	0.882	0.889*	0.845*	0.864	0.843*	0.862
	(0.078)	(0.079)	(0.096)	(0.095)	(0.098)	(0.097)
>1 marriage	1.892***	1.844***	1.519***	1.474***	1.478***	1.427***
	(0.071)	(0.072)	(0.100)	(0.101)	(0.097)	(0.097)
Age diff.	1.012***	1.011***	1.016***	1.014***	1.016***	1.015***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Husband						
Divorced	1.269***	1.274***	1.355***	1.354***	1.366***	1.365***
	(0.073)	(0.073)	(0.086)	(0.085)	(0.087)	(0.087)
Has kids	1.181**	1.187**	1.226**	1.234**	1.230**	1.238**
	(0.078)	(0.071)	(0.085)	(0.085)	(0.087)	(0.087)
Std. dev.	NA	NA	0.776***	0.764***	0.829***	0.824***
			(0.096)	(0.095)	(0.069)	(0.069)
Correlation	NA	NA	NA	NA	-0.860	-0.864
					(20.476)	(20.286)
Log-L	-10531.7	10537.3	-10511.7	-10518.4	-13545.2	-13552.0
Number of obs.	4021	4021	4021	4021	4021	4021

Notes: Sample weights are used. Estimates reported as hazard ratios. A coefficient greater than one indicates an increase in the hazard of marital dissolution while a coefficient of smaller than one indicates a decrease in the hazard. Standard error in parentheses. *** 1% ** 5% *** 10% significant different from 1.

Table 2.11: Probit Models for Premarital Cohabitation with and without Heterogeneity. Dep. Variable: Indicator for Premarital Cohabitation for First Three Marriages

Variable	I	II	III
Constant	0.956***	1.323***	1.268***
	(0.149)	(0.222)	(0.208)
> 1 marriage	0.286***	0.360***	0.354***
	(0.059)	(0.074)	(0.074)
Previous cohabitation	0.128**	0.201**	0.191**
	(0.053)	(0.080)	(0.078)
Intact family background	-0.176***	-0.259***	-0.246***
	0.040	(0.060)	(0.057)
Black	0.113	0.146	0.141
	(0.094)	(0.135)	(0.129)
White	0.064	0.098	0.093
	(0.078)	(0.110)	(0.106)
No religion	0.288***	0.329***	0.321**
	(0.090)	(0.136)	(0.130)
Catholic	-0.140*	-0.226*	-0.214*
	(0.081)	(0.124)	(0.120)
Protestant	-0.286***	-0.454***	-0.428***
	(0.078)	(0.123)	(0.117)
No high school	0.092*	0.100	0.099
	(0.052)	(0.077)	(0.074)
High school	-0.014	-0.041	-0.037
	(0.043)	(0.063)	(0.061)
Wife has kids	0.637***	1.009***	0.951***
	(0.050)	(0.085)	(0.077)
Husband has kids	0.100	0.152*	0.145*
	(0.062)	(0.082)	(0.080)
Husband is divorced	0.380***	0.496***	0.479***
	(0.060)	(0.084)	(0.080)
Not born in US	-0.390***	-0.535***	-0.512***
	(0.059)	(0.087)	(0.082)
Age at interview	-0.028***	-0.037***	-0.036***
	(0.003)	(0.005)	(0.004)
Standard Deviation	NA	0.916***	0.829***
		(0.098)	(0.069)
Correlation		NA	-0.860
			20.476
Log-Likelihood	-3060.2	-3033.0	-13545.2
Observations	4021	4021	4021

Notes: Sample weights are used. Standard error in parentheses. *** 1% ** 5% *** 10% significance levels

Table 2.12: Weighted Logit Estimator of Fixed Effects Model (Lee, 2003). Dep. Variable: Indicator 'Second marriage is longer than first marriage'

Variable	I	II	III
Premarital cohabitation	0.550**	0.687***	0.539**
	(0.269)	(0.260)	(0.252)
Trend	1.619***	1.877***	
	(0.390)	(0.365)	
Age at wedding (wife)	0.047	-0.300***	0.497***
	(0.211)	(0.047)	(0.186)
Age at wedding squared (wife)	-0.007*		-0.013***
	(0.004)		(0.004)
Age at wedding (husband)	0.026	0.018	0.065**
	(0.028)	(0.028)	(0.026)
Uncensored obs.	228	228	228

Notes: All 4030 observations were used to estimate base weights. A coefficient of greater than zero indicates that the variable increases marital stability, while a coefficient smaller than zero indicates that the variable decreases marital stability.
Standard errors in parentheses. *** 1% ** 5% * 10% significance levels.

Chapter 3

The Effect of Teenage Childbirth on High School Completion

Early childbearing is closely associated with adverse economic outcomes for mothers and their children such as poverty, welfare recipiency, lower educational achievement, and depressed earnings. It is tempting to describe these outcomes as consequences of early fertility decisions, but researchers have long ago recognized the problem with this conclusion. Teenage mothers are not a random sample of the population and the potential endogeneity of early childbearing can result in the close statistical association. One popular approach to overcome this difficulty is using instrumental variables which often take the form of 'natural experiments.' In this essay, I focus on the effect of adolescent childbearing on high school completion where depending on the choice of the instruments widely differing coefficient estimates emerge. I discuss the three main reasons why different instruments can result in conflicting estimates. By using two alternative instruments, I assess empirically which is the most likely in this particular application. One explanation focuses on potential defects of instruments which render the estimates unreliable. The second explanation posits that under treatment effect heterogeneity different instruments identify coefficients for different subpopulations as defined by their probability of experiencing teenage childbirth. Third, the different instruments may have different mechanism by which they affect the endogenous variable and this mechanism is in itself an important factor leading to differing coefficient estimates.

There is now a large literature assessing the socioeconomic consequences of adolescent childbearing, but here I want to focus on two studies which use instrumental variables. Ribar (1994) uses the National Longitudinal Survey of Youth (NLSY) to study the effect of teenage fertility on high school completion. He uses three different instruments: the availability of obstetricians and gynecologists, local abortion rates, and age at menarche. In his dataset he finds that teenage mothers have a 35 percent lower chance of finishing high school. After controlling for other observable characteristics teenage mothers have a 23.4 percent lower chance of finishing high school. Using age at menarche and the availability of obstetricians and gynecologists individually the effect of teenage childbearing on educational outcomes becomes even more negative (71.1 percent and 39.8 percent reduction in high school completion rates, respectively). However, if the local abortion rate is used as the sole instrument, the estimated coefficient turns sign. Using all three instruments together he finds a positive effect on high school completion rates, but this coefficient is not significant. Using a bivariate probit model he also finds strong indication for self-selection. Women who have high probabilities of early childbearing have unobserved characteristics that makes them more likely to drop out of high school. Hotz et al. (1997, 1999) use the NLSY and the occurrence of a miscarriage as an instrument and employ 2SLS regression models to estimate the effect. Their coefficient estimate on teenage childbirth indicates a positive effect of early childbearing on high school completion.

One should note that the studies by Ribar (1994) and Hotz et al (1997, 1999) use different sample definitions. Ribar includes all women in his sample while Hotz et al. restrict their sample to women who have become pregnant as teenagers. Ribar uses the whole sample because he has a variable that influences the probability that a women becomes pregnant. Hotz et al., on the other hand, restrict their sample to teenagers who have become pregnant before their 18th birthday. They interpret the estimated coefficient as a causal effect of teenage childbearing only for this subsample. They argue that one cannot use the occurrence of a miscarriage as an instrument for the whole population because miscarriages are correlated with pregnancies, and pregnancies are not a random event in the population as a whole. However, to compare the coefficient estimates using both instruments, one needs to use comparable samples. For this reason, I present results for different samples and assess whether these qualitative differences in instrumental variable estimates are robust to different sample definitions.

Conventionally, the coefficient on teenage childbearing in a 2SLS regression is interpreted as the causal effect of teenage childbearing on the outcome of interest.

Taking this view, it is difficult to reconcile the widely differing coefficient estimates. Maintaining the assumption of treatment effect homogeneity, there are two main potential problems with the instruments causing this problem. First, the instruments can be invalid. The assumption behind instrumental variable estimates is that the instruments only affect the outcome of interest via the endogenous regressor and that one can therefore safely exclude the instrument from the second stage regression. If one has more than one instrument, one can test these overidentifying restrictions. If there is only one fixed coefficient on the endogenous regressor, all instruments should result in estimates that are very close to each other, and one cannot reject the overidentifying restrictions. Instruments with weak explanatory power pose another problem (Staiger and Stock 1997), and the resulting estimates may even be more biased than OLS estimates.

In the last couple of years, interest in models with heterogenous treatment effects (Bjorklund and Moffitt 1987; Heckman and Robb 1985) emerged. In this context, using different instruments naturally lead to differing coefficient estimates. Imbens and Angrist (1994) point out that under treatment effect heterogeneity instrumental variables identify the Local Average Treatment Effect (LATE), that is the effect on those individuals whose treatment status has been changed by virtue of the instrument. Instrument variable estimates differ because the instruments may affect different subpopulations who differ in their treatment effects and in their propensity to undergo treatment. Angrist and Imbens (1995) show that under treatment effect heterogeneity the coefficient estimate using all instruments together is a weighted average of the coefficient estimates using the instruments individually. Heckman, Urzua, and Vytlacil (2006) propose a test for treatment effect heterogeneity by looking for nonlinearities in the relationship between the probability of participation and the outcome of interest. In this view, different instrumental variable estimates are not necessarily a sign of their defects and simply may reflect treatment effect heterogeneity in the population. The instruments in the literature on teenage childbearing include contextual variables as local abortion rates and access to medical care as well as natural experiments such as miscarriages and age at menarche. It is reasonable to assume that the consequences of teenage childbearing may be heterogenous among the population and that the instruments affect different subpopulations as defined by their probability of experiencing a teenage childbirth. For example, a miscarriage will disproportionately affect teenagers who presumably have had a high probability of teenage childbearing while the age at menarche may affect teenagers with more moderate probabilities of teenage childbearing.

Finally, Moffitt (2005) discusses the problem of the multiplicity of mechanisms by which an instrument influences the endogenous regressor. In the case of teenage childbearing, women may have different reasons to postpone a birth (or not give birth at all) after teenage years. A high age at menarche and miscarriage both lead to a higher age of the individual at the birth of the first child and maybe to a postponement of having children after the women's teenage years. But the mechanism for this postponement is different. It may make a difference in the decision to drop out of high school whether an individual gives birth to a child at age 15 versus age

17. Using a binary indicator for teenage childbearing masks these differences. If one uses instruments that affect different populations as defined by their age at first pregnancy, then these differences in the population affected may explain differences in the estimates.

In this essay, I investigate empirically which of the three explanations is most likely in the case of the effect of teenage childbearing on high school completion. The essay is organized as follows. After discussing the dataset and the determinants of early childbearing, I estimate 2SLS coefficients on early childbearing where the dependent variable is an indicator for not finishing high school using the instruments individually and jointly. In a first step, I investigate to which extent defects in the instruments drive the differences in coefficient estimates. Having more than one instrument allows me to test the overidentifying restrictions in addition to the explanatory power of the instruments in a first-stage regression. I also investigate whether my results are robust to different sample definitions. Using different sample definitions also help to compare my results to Ribar (1994) and Hotz et al. (1997, 1999). Then, I explicitly model treatment effect heterogeneity using a model by Moffitt (2007) allowing to directly test for treatment effect heterogeneity where the treatment effect depends on the probability of early childbirth. Furthermore, I estimate marginal treatment effects to assess whether this kind of treatment effect heterogeneity is likely to explain the differing coefficient estimates when using alternative sets of instruments. Finally, I replace the binary indicator for a teenage childbirth with a more detailed variable taking into account the age at first conception and investigate whether the two instruments induce differences in the distribution of age at first birth. If age effects are important, then this can also explain differences in the instrumental estimates.

3.1 Data and Empirical Models

3.1.1 Descriptive Statistics

The sixth circle of the National Survey of Family Growth (NSFG) is a representative sample of 7643 women aging from 15 to 44 years at the date of survey in 2002. Women were asked about important life events including relationships and pregnancies. I restricted my sample to women in the age range 20 to 44 in order to ensure that the individuals have had an opportunity to finish high school. This cut-off was based on Hotz et al. (1997) who discuss that high school completion rates do not change substantially after age 20. My resulting sample consists of 6443 women reflecting mainly the smaller age range than the original sample. I measure high school completion by the number of completed years of grades. If the woman has not successfully completed twelfth grade, the dummy for less than high school takes the value 1[1]. In addition, the interviewers asked about family background. From these variables, I constructed dummies for the mother's and father's educational achievement, whether the parents have separated or are divorced, and whether the father

[1]It should be pointed out that in the group of women who did not complete twelfth grade, teenage mothers are more likely to obtain a GED than other women. However, for males there is evidence that high school graduation and the GED are not equivalents (Cameron and Heckman 1993).

was present. For a small number of cases educational achievement was not known and this category serves as omitted category. Another familial background variable is a dummy for whether the individual was born in the US or abroad. Furthermore, women were asked about their religion and their race. Following the NSFG recodes, I constructed dummies for Blacks and Whites where other races are the omitted category. The religious dummies consist of one for Catholics, Protestants, other religions, and non-religious.

Teenage mothers have different characteristics than other teenagers. I present some descriptive statistics in table 3.1 for teenagers with a teenage birth ($D = 1$) and without a teenage birth ($D = 0$). Of teenage mothers around 51% do not have a high school degree at the time of the interview while only 14% of women without a teenage childbirth do not have a high school degree. This raw difference is very similar to Ribar (1994) who found a 35% difference in high school completion rates between these two groups. Teenage mothers are slightly younger than other women indicating a rise in the proportion of teenage mothers over the last decades. Teenage mothers are also more often black. But in terms of religious affiliation, there does not seem to be a big difference. Teenage mothers are also more likely to come from families in which the parents have separated, and where the parents have lower educational achievement. These summary statistics all point to the well known fact that teenage mothers often come from economically more disadvantaged backgrounds.

Following Ribar (1994), I use age at menarche as one of my instruments. In the survey, participating women were asked at which age (in years) they had their first menstrual period. The variable age at menarche is based on this variable. For a very small number of women in the survey the menstrual period has not started yet, or they don't know or refuse to answer. These observations were dropped. In addition to this variable, I also a constructed a binary dummy for an early menstrual period. This dummy takes the value 1 if the woman was less than 13 years at the time of their first menstrual period (13 years is also the median age at menarche in this sample). The distribution is fairly symmetric around this median age, and around half of all observations are at age 12 and 13. Following Hotz et al. (1997, 1999), I construct a dummy for a miscarriage in the first pregnancy. In the survey, participating women were also asked about their pregnancies and pregnancy outcomes. The dummy for a miscarriage takes the value 1 if the woman reported a miscarriage (655 cases) during the first pregnancy. For a smaller number of women (89), I also assigned the value 1 for an ectopic pregnancy or a stillbirth. The dummy for having a teenage childbirth takes the value 1 if the baby is a lifebirth and the date of conception is before the mother's 18th birthday[2]. In table 3.1 I present the means for these instruments for the group of women who have experienced a teenage childbirth and for the group who have not. Teenage mothers were younger at the age of their first menstrual period and they are less likely to have experienced a miscarriage in their first pregnancy.

Of separate interest is the question whether some of the socioeconomic variables are good predictors of the occurrence of a miscarriage and early menarche. This may

[2]This definition of a teenage child birth follows Hotz et al. (1999) who require that the woman conceives before her 18th birthday.

point to differences in observable characteristics in the subpopulations from which individuals with specific instrument values are drawn. For this reason, I present OLS regressions of indicators for the occurrence of a miscarriage, age at menarche, and a dummy for early menarche on the other socioeconomic variables on which I condition throughout my analysis (table 3.2). In the first column of these tables, one finds the results with the occurrence of a miscarriage as the dependent variable. Two of the socioeconomic variables have a big influence on this variable. The older the individual at the age of the interview, the more likely it is that she has experienced a miscarriage during the first pregnancy. Of course, the older the individual is, the higher are the chances that the woman has ever been pregnant. The other significant variable is whether the woman has an intact family background. Women whose parents have separated are more likely to have experienced a miscarriage. This association arises because women whose parents separated are also more likely to have been pregnant early. Because of this I also run the same regression of the occurrence of a miscarriage on all socioeconomic variables for women who have ever been pregnant in their lives (column 2). I find that none of these variables is significant for this subsample. For the whole sample there is a connection because age and family background are associated with the risk of becoming pregnant.

Like in most of the literature, I initially focus on a binary indicator for teenage childbearing. However, it should be noted that miscarriages seem to disproportionately affect mothers who have conceived before their 15th birthday as one can see third column of table 3.2 [3]. For women younger than 15 years old at the beginning of their first pregnancy, there is a significantly increased risk of a miscarriage. There is also medical evidence that a young age at the start of the pregnancy is a risk factor for miscarriages (Fraser, Brockert, and Ward 1995). On the other hand, early menarche makes an early conception more likely because the woman starts to be at risk for a pregnancy at a younger age.

In the fourth column, I present the regression with a dummy for early menarche as independent variable. Two variables are highly significant. There is a strong age effect. Older women at the time of the interview are less likely to have experienced an early menarche. This time trend in the fall of the age at menarche has also been noted in the medical literature (Freedman, Khan, Serdula, Dietz, Srinivasan, and Berenson 2002). This latter study also finds that black girls have a lower age at first menarche than other races. However, racial variables are insignificant in my analysis. The other significant variable is whether the woman was born in the United States. Women born in the US have a lower age at menarche. One might think that different nutritional levels in childhood could play a role here (Freedman, Khan, Serdula, Dietz, Srinivasan, and Berenson (2002) also point to nutrition as one potential factor). A similar pattern emerges if age at menarche (column 5) is the dependent variable. Surprisingly, here the father's education seems to play an important role with a higher age at menarche for women whose father has at least a college degree.

[3] See also the later discussion and figure 3.9.

Thus, one can conclude that the instruments affect different subpopulations of women as defined by some observable variables. In particular, a very young age at the beginning of the first pregnancy makes a miscarriage more likely. This results in differences in the age distribution at first pregnancy associated with each of the instruments[4]. In addition, domestic born and younger women have a lower age at menarche.

3.1.2 The Determinants of Early Child Bearing

Socioeconomic Factors of Early Child Bearing

The determinants of early childbearing may be of independent interest to the researcher. Furthermore, one concern with instrumental variable estimation is that the instruments are weak in the first-stage regression (Staiger and Stock 1997). Weak instruments have low explanatory power in the first-stage regression, and their use may lead to biased estimates. For this reason, I present OLS regressions where the dependent variable is a dummy for a teenage childbirth and probit regressions. The OLS regression results serve as first-stage regressions for the 2SLS estimates. The probit regressions will later be estimated jointly with the outcome equation when estimating Moffitt's (2007) model with treatment effect heterogeneity. I compared those latter results to the simple probit regressions and found that there are small quantitative differences, but the estimates are otherwise qualitatively similar. For this

[4]See also figures 3.8 and 3.9.

reason, I present only the simple probit regressions. I will also graph the predicted probabilities for having a teenage childbirth using both instruments. This is important because nonparametrically, the marginal treatment effects are only identified within the support of the participation probabilities.

The determinants of the decision to bear a child as a teenager are presented in table 3.3 (OLS) and table 3.4 (Probit). Both models give qualitatively similar results. I estimated five different specifications using the instruments both individually and jointly. The inclusion of different sets of instruments does not seem to alter the other coefficient estimates. In all specifications the coefficients on the instruments take the expected sign. Having a miscarriage in the first pregnancy dramatically reduces the probability of having a teenage childbirth. An early menarche both measured as a dummy or as a continuous variable increases the probability of a teenage childbirth. For the OLS model, I also conducted F-tests for the exclusion of the instruments from the equation (see table 3.5). All instruments have very high explanatory power, and therefore I conclude that there does not seem to be a weak instruments problem (see Stock and Yogo (2002) for critical values).

Older women at the time of the interview have a lower probability of having had a teenage childbirth in their youth. Race and ethnicity play an important role. Blacks have a higher probability than other groups, and foreign born women have a lower probability of having a teenage childbirth. Perhaps surprisingly, religious affiliation does not seem to play an important role. I should note, however, that this is measured at the time of the interview, and therefore might differ from the religious affiliation at the time the women was a teenager. An important factor is family background. Women who grew up in intact families have lower probabilities of experiencing a teenage childbirth than women whose parents separated or where the father was missing for other reasons. Furthermore, parents' education is important. The better educated the parents, the lower are the probabilities that the child becomes pregnant as a teenager.

I also present kernel estimates of predicted probabilities of having a teenage pregnancy both for the OLS and the Probit (figure 3.1) model. For the OLS model, the range of predicted probabilities is from about -0.16 to 0.62 with a mean of 0.14 and standard deviation of 0.10. The negative predicted probabilities will pose no problem for the calculation of the 2SLS estimates. However, they are problematic for calculating the marginal treatment effects in Moffitt's model since here it would be difficult to interpret observations for which the predicted probability is less than zero. The probit model in figure 3.1 forces the predicted probabilities to be positive. The range of predicted probabilities is between 0.00 and 0.78 with a mean of 0.14 and standard deviation of 0.11. The distribution looks more skewed than the corresponding OLS estimates because the mean of the predicted probabilities is relatively close to zero and there cannot be negative predicted probabilities. For both models a lot of probability mass is concentrated at low probabilities and therefore my estimates for the marginal treatment effects using Moffitt's will be most reliable in this area. Marginal treatment effects for very high participation probabilities will rely solely on extrapolation. The average treatment effect cannot be estimated non-parametrically since there are no women for which the predicted probability of an early childbirth is close to one.

Different instruments can result in different estimates because they may affect different subpopulations who have different probabilities of having a teenage childbirth. For this reason, it is important to assess how the instruments affect different subpopulations. Figure 3.2 presents kernel estimates of the predicted probabilities of having a teenage childbirth using a probit model and the occurrence of a miscarriage and a dummy for early menarche as instruments. The figure contains these estimates for the whole sample and the predicted probabilities for early childbearing if the instruments are set to alternative values. For women who experience a miscarriage more probability mass is shifted to the left which is not too surprising since it reflects the dramatic reduction in probability of having a teenage childbirth after experiencing a miscarriage in the first pregnancy. For women with an early menarche more probability mass is shifted to the right indicating an increase in the probability of having a teenage childbirth. The difference between the instruments is that miscarriage has a quantitatively stronger effect than age at menarche. This can also be seen in figure 3.3 where box plots at four quartiles of the predicted probabilities are shown. Again, the instruments are set to distinctive values. Experiencing a miscarriage has a quantitatively much more important effect on these predicted probabilities than changes in the menarche variables. Furthermore, the impact of all instruments on the predicted probabilities seems to be strongest in the higher quartiles.

3.1.3 2SLS Estimates with and without Heterogeneity

A 2SLS estimate can be given two different interpretations depending on whether one assumes the presence of homogenous or heterogenous treatment effects where the treatment effect depends on the probability of experiencing an adolescent childbirth. The 2 Stage Least Squares under treatment effect homogeneity estimator takes the following form

$$Y_i = X_i\beta + \gamma_0 * D_i + e_i \tag{3.1}$$

D_i is the potentially endogenous variable, teenage childbirth, with the property that $E(D_i e_i) \neq 0$. Instruments can help to deal with this endogeneity. An instrument, Z, has to satisfy the following two assumptions:

$$E\left(Z'D\right) \neq 0 \text{ (instrument relevance)} \tag{3.2}$$

$$E\left(Z'e\right) = 0 \text{ (instrument exogeneity)} \tag{3.3}$$

The first assumption states that the instrument is correlated with the potentially endogenous variable while the second assumption states that the instruments is not related to the error term e_i. The standard IV estimator can then be obtained by regressing the endogenous variable on all exogenous variables (X_i) and instruments (Z_i) excluded from X_i. The first stage regression is then given by:

$$D_i = X_i\delta + Z_i\eta + u_i \tag{3.4}$$

The predicted $\widehat{D_i}$ from this first stage regression is plugged into the second stage regression.

$$Y_i = X_i\beta + \gamma_0 * \widehat{D_i} + e_i' \tag{3.5}$$

By construction $\widehat{D_i}$ is not correlated with the error term e_i'. If one assumes treatment effect homogeneity, the coefficient $\gamma_0 = \Delta^{MTE} = \Delta^{ATE} = \Delta^{ATT} = LATE$. These parameters refer to the marginal treatment effect, the average treatment effect, the average treatment effect on the treated, and the local treatment effect (see Heckman and Vytlacil, 2005 for a discussion of these different effects.) Under treatment effect heterogeneity these parameters do not need to be the same. In this case, the marginal treatment effect depends on the probability that $D_i = 1$. Imbens and Angrist (1994) discuss the case of binary instruments. In this case, the instruments identify the local average treatment effect ($LATE$) for the individuals who are moved from one treatment state to the other by virtue of the instrument. The instruments may affect two distinct groups for which the $LATE$ needs not be the same, making it possible that one instrument identifies the $LATE^1 = \gamma_0^1$ for the first group and the other identifies $LATE^2 = \gamma_0^2$ for the second group of individuals.

If there are more than one instrument, one can test the overidentifying restrictions, that is whether one can safely exclude the instruments from the second stage regression. If both instruments are valid, and there are homogenous treatment effects, the estimates using different instruments should not be too different, and one should not reject the overidentifying restrictions. This needs not be the case under treatment effect heterogeneity. If one writes down the moment conditions for each instrument, one finds why one might reject the overidentifying restrictions, even though the instruments are valid:

$$E(Z_1(Y - X\beta - \gamma_0^1 * D)) = 0 \tag{3.6}$$

$$E(Z_2(Y - X\beta - \gamma_0^2 * D)) = 0 \tag{3.7}$$

If the instruments are exogenous, and there is treatment effect homogeneity ($\gamma_0^1 = \gamma_0^2 = \gamma_0$), both moment restrictions can be satisfied at the same time. However, consider the case of treatment effect heterogeneity, that is $\gamma_0^1 \neq \gamma_0^2$, and each instrument identifies a different parameter. Estimating the 2SLS model one forces the coefficient estimates to be equal ($\widehat{\gamma_0^1} = \widehat{\gamma_0^2} = \widehat{\gamma_0}$). Then it is clear, that one cannot satisfy both moment restrictions at the same time, and one would reject them even though the instruments are valid.

Another specification test is to check whether the instruments are weak. Weak instruments do not have a lot of explanatory power in the first-stage regression and result in biased estimates. Stock and Yogo (2002) suggest computing F-tests in the first-stage regression for joint significance of the instruments and provide critical values to assess whether weak instruments are a problem.

There is one further complication in this case. The variable Y_i is a dummy for not finishing high school and not a continuous variable. This introduces the problem of heteroscedasticity in a linear probability model (Aldrich and Nelson 1984). Furthermore, the Sargan test-statistic of the overidentifying restrictions is only valid with homoscedastic errors. For this reason, I also estimate the feasible efficient two-step

GMM estimator. The idea is to estimate in a first step the residuals $\widehat{e_i}$ using the outlined 2SLS estimator and use those estimates to derive an optimal weighting matrix for the second step. The overidentifying restrictions can then be tested with Hansen's J statistic.

The estimated coefficient on having a teenage childbirth are presented in table 3.5. The OLS estimate shows a strong and statistical significant association between having a teenage childbirth and not finishing school. Teenage mothers face a 29.3% higher probability of not finishing high school according to these estimates. The estimate of the coefficient using a miscarriage shows a 41.8% reduction in this probability. This can be interpreted as indicative of self-selection of mothers with low prospects of finishing high school into early motherhood. When one is using age at menarche as an instrument, the coefficient indicates that teenage mothers face a 14.7% higher probability of not finishing high school. This estimate is below the OLS estimate, and it also indicates self-selection of mothers with low prospects of finishing school into early motherhood. When one uses a dummy for early menarche, the estimated coefficient is about the size of the OLS estimate. If one uses the occurrence of a miscarriage together with age at menarche or the dummy for an early menarche, one finds coefficient estimates on early childbearing of about -0.13 corresponding to a 13% reduction in risk of dropping out of high school. Adding the square of age at menarche does not change these coefficient estimates much. This point estimate is consistent with Angrist and Imbens' (1995) result which showed that the coefficient estimates using more than one instrument is a weighted average of coefficient estimates when using the instruments individually.

In table 3.5, I also present diagnostic checks for whether the instruments are valid and relevant. I report both Sargan's and Hansen's J statistic. Using miscarriage together with either age at menarche or the dummy for early menarche, I can reject the overidentifying restrictions on the 5% level. On the other hand, I cannot reject the overidentifying restriction using age at menarche and its square. However, this test is hard to interpret because of the high collinearity between the two instruments. Finally, using miscarriage, age at menarche, and its square, I cannot reject the overidentifying restrictions using Hansen's J statistic. These results point to potential problems with the validity of the instruments, especially of miscarriage.

Further tests of the explanatory power of the instruments show that instruments are not weak. The F-statistic of the test of joint significance of the instruments in the first stage regression show very high values for all instruments (see Stock and Yogo (2002) for critical values).

I also investigated the question whether the estimated effects are different for the subsample of women who became pregnant before they reached the age 18 as Hotz et al. (1997, 1999) have done. For the whole population, miscarriage may fail as an instrument because pregnancies are not randomly distributed in the population and miscarriage is correlated with it. There are 1284 women in the teenage pregnancy sample. The first panel of table 3.6 presents these results. The OLS coefficient on the dummy for a teenage childbirth is still significant but reduced in comparison to the whole sample. This points to self-selection. Comparing women who had a teenage childbirth to other teenagers who became pregnant but did not have a life birth reduces the statistical association between teenage childbirth and educational achievement. When using the occurrence of a miscarriage one finds a beneficial effect of having a teenage childbirth on educational outcomes, but this effect is smaller than for the whole sample. Employing early menarche as an instrument one finds an even stronger adverse effect of early childbirth on education achievement. However, both age at menarche and early menarche are weak instruments in this subsample. This is not too surprising since the effect of an early menarche presumably increases the risk of young women to become pregnant. If one includes only pregnant women in the sample, then there should be only a weak relationship between age at menarche and experiencing a teenage childbirth. Even though there is a weak instrument problem in this subsample now, there does not seem to be a problem with the validity of the instruments. In the second panel of table 3.6, I present results for a sample of women who have ever been pregnant. There are 4810 observations in this sample. Both instruments should perform well: Since all women have been pregnant, there is no correlation between experiencing a miscarriage and pregnancies. Age at menarche does not need to be a weak instrument in this sample because it affects the exact timing of the birth – whether as a teenager or later. The OLS estimate of the effect of early childbearing is again of a similar magnitude than in the other samples. Again, the coefficient estimates based on miscarriage and age at menarche are of opposite sign, but the quantitative difference is not as pronounced as in the other samples. No instrument suffers from weakness, and one cannot reject the overidentifying restrictions, when one uses age at menarche and the occurrence of a miscarriage as instruments. Finally, in the third panel of table 3.6, I present results for a sample of women who have experienced a lifebirth resulting in 4273 observations. For similar reasons as before, all instruments should perform well. They are not weak, and one cannot reject the overidentifying restrictions. In this subsample, the same qualitative picture emerges: using miscarriage as an instrument, one finds a beneficial effect of teenage childbearing on high school completion while using menarche, one finds an adverse effect.

These results can be compared to Hotz et al. (1999) and Ribar (1994) whose estimates are presented in table 3.7. My OLS estimate of the effect of teenage childbearing lies between the Ribar and the Hotz et al. estimates. When Hotz et al. (1999) use high school completion as the outcome and miscarriage as an instrument in their teenage pregnancy sample, the coefficient on teenage childbearing is greatly reduced but still positive. Overall teenage pregnancy would still reduce high school completion rates. However, if he uses high school completion and GED, he finds that teenage pregnancy improves educational achievement qualitatively consistent with my estimates of the effect of teenage child bearing. My estimates, however, indicate a stronger effect than Hotz et al. (1999). Ribar (1994) finds that when he uses only age at menarche as an instrument for teenage childbearing, the effect on educational achievement is even more detrimental. According to this estimate, teenage childbearing reduces the probability of high school completion by 71.1% which is even higher than the corresponding OLS estimate. Overall, my results are broadly in line with their results: When using miscarriage as an instrument, one finds that teenage childbearing has a very small and possibly even positive effect on educational achievement. On the other hand, when using age at menarche as an instrument, one finds that a teenage birth reduces educational achievement. Unfortunately, I could not use the other contextual variables in Ribar (1994) because the public use NSFG 2002 lacked geographical identifiers.

3.1.4 A Model for Heterogenous Treatment Effects

In the presence of heterogenous treatment effects, Heckman and Vytlacil (2005) propose estimating marginal treatment effects (MTE's) to analyze policy questions. The MTE is the effect on a person who is brought into the program by a marginal extension of the program under question. Obviously, a teenage pregnancy is not a policy program like a job market training program. Nonetheless, people are interested in the effects of a teenage pregnancy on some economic outcomes like educational achievement, poverty status, income, etc. Furthermore, there are policies that might be thought of as encouraging or discouraging teenage pregnancies. Such policies include welfare and tax policies or the availability of health care for young mothers. Given the MTE, one can then analyze these policies affecting the likelihood of teenage childbirth and the subsequent consequences on educational achievement or other outcomes of interest.

Moffitt's Model

Moffitt (2007) proposes a model to estimate marginal treatment effects in heterogenous populations. This model is closely related to the 2SLS estimates, but it allows for a treatment effect heterogeneity through nonlinearities in the function relating the participation probabilities to the outcomes. Treatment effect heterogeneity may take the following form:

$$
\begin{aligned}
Y_i &= h(X_i) + \gamma_i * D_i + \varepsilon_i && (3.8)\\
D_i^* &= k(X_i, Z_i, \delta, \eta) && (3.9)\\
D_i &= \begin{cases} 1 \text{ if } D_i^* > 0 \\ 0 \text{ otherwise} \end{cases} && (3.10)
\end{aligned}
$$

To allow better comparability to the 2SLS model, I will impose linearity in the X_i such that $h(X) = X\beta$. The crucial difference to equation 3.1 is that γ_i is a random coefficient which may take on different values for different individuals.

Conditioning on X_i and Z_i in 3.9 , 3.10, and 3.10, one gets:

$$
\begin{aligned}
E\left(Y_i|X_i, Z_i\right) &= X_i\beta + E\left(\gamma_i|X_i, Z_i, D_i = 1\right) * \Pr\left(D_i = 1|X_i, Z_i\right) && (3.11)\\
E\left(D_i|X_i, Z_i\right) &= \Pr\left(D_i = 1|X_i, Z_i\right) && (3.12)
\end{aligned}
$$

The following assumption on the expectation of the treatment effect is made:

$$E\left(\gamma_i|X_i, Z_i, D_i = 1\right) = g\left(E\left(D_i|X_i, Z_i\right), X_i\right) \tag{3.13}$$

The instruments are valid if the following assumption holds:

$$E\left(\varepsilon|X, Z\right) = 0 \tag{3.14}$$

This assumption ensures that Z_i satisfies the exclusion restriction. It says that the Z_i influence the treatment effect only through their effect on the participation probability but not directly. It allows also full interaction between the X_i and the treatment effect. One needs also to impose a 'monotonicity' assumption (see Imbens and Angrist 1994)

$$D_i\left(X_i, Z_i = z\right) - D_i\left(X_i, Z_i = z'\right) \tag{3.15}$$

is zero or the same sign for all i for any distinctive value z and z'.

Let $F\left(X_i, Z_i\right) = \Pr\left(D_i = 1 | X_i, Z_i\right)$. The model can then be written as

$$Y_i = h(X) + g\left(F\left(X_i, Z_i\right), X_i\right) * F\left(X_i, Z_i\right) + \varepsilon_i \tag{3.16}$$

$$D_i = F\left(X_i, Z_i\right) + \nu_i \tag{3.17}$$

By construction the errors are mean-independent from the regressors. The $g()$ function can be a nonlinear function. For homogenous treatment effects $g()$ is a constant. Treatment effect heterogeneity implies nonlinearities in the relationship between Y_i and $F\left(X_i, Z_i\right)$. In my case, I have shown that the distributions of the participation probabilities conditional on different combinations of the instruments are different.

The following is a specialization of the more general model which I use for my empirical analysis. The decision to have an early childbirth is modeled as a probit model, such that $\Pr\left(D_i = 1 | X_i, Z_i\right) = \Phi\left(X_i\delta + Z_i\eta\right)$ where $\Phi()$ is the cdf of the standard Normal distribution. Furthermore, I assume that there are no interactions between the $X_i's$ and the treatment effect[5] and that the relationship between the participation probability and the treatment effect is quadratic such that $g\left(\Phi\left(X_i\delta + Z_i\eta\right), X_i\right) = \gamma_0 * \Phi\left(X_i\delta + Z_i\eta\right) + \gamma_1 * \left[\Phi\left(X_i\delta + Z_i\eta\right)\right]^2$. The model is then given by:

$$Y_i = X_i\beta + \gamma_0 * \Phi\left(X_i\delta + Z_i\eta\right) + \gamma_1 * \left[\Phi\left(X_i\delta + Z_i\eta\right)\right]^2 + \epsilon_i \tag{3.18}$$

$$\Pr\left(D_i = 1|X_i, Z_i\right) = \Phi\left(X_i\delta + Z_i\eta\right) + \nu_i \tag{3.19}$$

There are a couple of differences to the 2SLS regression. First, in the 2SLS regression, the coefficient γ_1 would be zero, since there is only a linear term in this model. Second, in the 2SLS model the first stage is also an OLS linear probability model. I chose to use a probit specification because this restricts the estimated participation probabilities to be nonnegative whereas in the OLS specification there were some negative predicted participation probabilities. Notice that given the functional form assumptions, one can calculate the Δ^{MTE} for all participation probabilities, even those outside of the support of $\Pr(D_i = 1|X_i, Z_i)$ They can be calculated as the partial derivative $\frac{\partial Y_i}{\partial P_i} = \gamma_0 + 2 * \gamma_1$. A test for treatment effect heterogeneity would be a test of the restriction $\gamma_1 = 0$. Third, the model will be estimated jointly by Generalized Nonlinear Least Squares. Details can be found in Moffitt (2007). With a binary instrument in the Moffitt model, one can only identify the marginal treatment effect between two points induced by the two values of the instrument. In this application,

[5]I also estimated a more general model where I allow for this sort of interaction. These interaction effects have very large standard errors and are statistically insignificant.

however, I am interested in the overall shape of this function. Identifying this shape requires a continuous instrument to map out the function linking the participation probability to the marginal treatment effects. For this reason, all my estimates of the Moffitt model include age at menarche which is approximately a continuous variable.[6] In addition, including the square of age at menarche helps in estimating the overall shape of the marginal treatment effect function because it induces a wider spread in the predicted probabilities of teenage childbirth. Identifying the average treatment effect would require an instrument that induces predicted probabilities of close to 1. In this application, however, most predicted probabilities are below 0.4. Therefore, it is not possible to estimate the average treatment effect using Moffitt's model. Since the focus in this part is on treatment effect heterogeneity, I focus on the whole sample. If there is treatment effect heterogeneity, one is more likely to find it in a more heterogenous population. Therefore, I do not restrict my sample here to, for example, only women who have become pregnant as teenagers.

In a first step, I estimate the Moffitt's model without nonlinearities in the relationship between the participation probability and the outcome by setting $\gamma_1 = 0$. Table 3.8 presents these results. Thus, the estimated treatment effects should correspond to the 2SLS regression estimates using the same instruments. The differences between the estimates using Moffitt's model compared to the 2SLS model are generally small. The biggest difference is in the coefficient on teenage childbearing when the only in-

[6] Strictly speaking, of course, it takes on discrete values.

strument is age at menarche. The Moffitt model then predicts an about 10% higher chance of not completing high school than the corresponding 2SLS estimate, but the standard errors are wide.

The results for the Moffitt model with treatment effect heterogeneity are displayed in table 3.9. Since treatment effect heterogeneity takes the form of nonlinearities of the function linking participation probabilities to marginal treatment effects the coefficient γ_1 shows the degree of treatment effect heterogeneity in the effect of having a teenage childbirth on educational achievement. The coefficients on the participation probability differ widely depending on which instruments are used and the standard errors are rather big. All coefficients on γ_1 are negative which implies that the function linking the participation probability with the outcome is concave. At higher probabilities of having a teenage pregnancy a teenage childbirth affects educational outcomes less adversely or might even be beneficial. Notice however, that the coefficients on γ_1 are not significant so that these results are consistent with the view that there is treatment effect homogeneity. Of course, this lack of significance could arise if the true nonlinear function is badly approximated by a quadratic function.

Figures 3.4 to 3.7 show the function linking the marginal treatment effects to the probability of early childbirth using alternative sets of instruments. The function to the right of about 0.4 are an extrapolation because there are almost no observations with these predicted probabilities. In figure 3.4, I use only the age at menarche as an instrument. The adverse effect of teenage childbearing on educational achievement is more adverse at low predicted probabilities of a teenage childbirth, but the standard errors are very wide and include the zero for a wide range of predicted participation probabilities. If one uses miscarriage and age at menarche as instruments (in figure 3.5), one finds almost no indication of treatment effect heterogeneity. There is a small positive effect on educational achievement, but it is not statistically different from zero. If one uses age at menarche and its square as instruments (figure 3.6), there are no noticeable differences to the graph where only age at menarche is used. Finally, for the model where miscarriage, age at menarche, and its square are used, there are again no big differences to the results in figure 3.7. In none of the models using the different sets of instruments does one find strong indication for treatment effect heterogeneity. If this is the case, then differences in the estimates using alternative sets of instruments are unlikely to be explained by treatment effect heterogeneity. The estimated coefficient on teenage childbearing in the 2SLS estimates is positive when using age at menarche and negative when using miscarriage. If treatment effect heterogeneity were to explain these differences, the function linking marginal treatment effects to the participation probability would have to be positive for one range of probabilities and negative for a different range. In addition, the instruments would have to sweep out exactly these ranges to result in 2SLS estimates switching their sign. This seems to be rather unlikely.

The other coefficients do not change much depending on which instruments are used. Foreign born women have a harder time finishing their high school education. The most significant factor is family background. Women coming from non-intact families or where the father was absent have lower probability of finishing high school.

The better educated the parents are, the higher are the chances that the individual finishes school. Religious background seems to play a role. Perhaps surprisingly, race does not seem to play a large role in determining educational outcomes after controlling for other background factors.

3.1.5 Is Using a Binary Regressor for Teenage Childbearing Justified?

Moffitt (2005) points to the importance of understanding the exact mechanism by which the instruments affect the endogenous regressor when conducting reduced form analysis as in this particular application. Both miscarriage and age at menarche affect the chance of having a childbirth as a teenager when this event is measured as a binary treatment variable. However, there are other plausible channels by which both instruments affect intervening variables that have been omitted from the analysis. Here, I want to focus on the age of the teenager when she gives birth. It may be plausible that giving birth at age 14 may have different effects on high school completion than giving birth at age 17. These differences are masked when using a binary indicator for having a teenage childbirth. If the effects are different by age group, and the instruments affect different populations as defined by their age,[7] this would be another explanation for differing instrumental variable estimates when us-

[7]This can be also seen as treatment effect heterogeneity, but in this case it depends on observable variables.

ing alternative sets of instruments. In this context, it is important to note that the mechanism by which the instruments affect the treatment are completely different. Age at menarche is associated with teenage pregnancy presumably because teenagers with an early menarche are at risk for getting pregnant at a younger age than other teenagers. On the other hand, miscarriages affect women who have already become pregnant. Both mechanism may result in differences in the age distribution at child-birth induced by the two instruments, especially since there is evidence that a very young age at first pregnancy is in itself a risk factor in miscarriages (see figure 3.9 and table 3.2).

I show kernel estimates of the ages at first conception for teenagers with early and late menarche (figure 3.8) and for teenagers who have and have not experienced a miscarriage during their first pregnancy (figure 3.9). Both subpopulations are compared with the whole population. For the age range of interest for this essay, it looks as if the age at menarche shifts the distribution of age at first conception in the way one would expect. Women with an early menarche have more probability mass on lower ages at first conception. This effect on age at first conception seems to be quite uniform across all age groups in the range that is of interest here, say before age 21. On the other hand, if one compares this picture to the kernel estimates for the subpopulations defined by the occurrence of a miscarriage (or no miscarriage), one finds that a miscarriage disproportionately affects women who became pregnant before their 15th birthday while it affects older teenagers considerably less. This fact has also been noted in Hotz et al. (1999). In addition, Fraser, Brockert, and Ward

(1995) find that younger teenagers are more likely to experience other poor pregnancy outcomes such as low birth weight. One may speculate that medical complications during pregnancy add additional stress and may be a factor in the decision to drop out of high school. Comparing the two graphs reveals that age at menarche and miscarriage affect different populations of teenage mothers as defined by their age at first conception. This would not be a problem for the analysis with a binary indicator for teenage childbearing if the effect is homogenous across different age groups. However, this is only an assumption, and one can test whether this is true.

In the analysis so far, the definition of having a teenage childbirth required a conception before the 18th birthday. I constructed additional dummy variables requiring conception before the 15th, 16th, and 17th birthday to uncover any age effects. Again, the omitted category is postponement or not giving birth at all. In addition, I constructed dummy variables for giving birth before age 15, ages 15-16, ages 16-17, ages 17-18, and ages 18-19. The omitted category is postponing birth after adolescence or not giving birth at all. In a first step, I used OLS estimates controlling for the covariates from the previous analysis and these new age categories. Table 3.10 presents these results. In the first four columns, one can find the results for the dummies indicating a teenage childbirth requiring conception before the 15th, 16th, 17th, and 18th birthday. The coefficient for very early childbirth with age of conception before the 15th is higher than all the other coefficients using this set of dummy variables. This can be more clearly seen by looking at the result in the fifth column. A teenage childbirth before the age 15 is associated with an increase in the probability of not finishing high school by 49% while this risk is reduced to 18.2% for teenagers giving birth after their 18th birthday.[8] Of course, age at first childbearing is probably again an endogenous variable and the same problems of self-selection are likely to be important here. Still, the results are suggestive of important heterogeneity across different age groups which are masked by using a binary indicator for teenage child bearing.

To deal with the endogeneity of these new dummy variables, I used again age at menarche as a dummy for the first set of dummy variables requiring different ages at conception for the first teenage childbirth. For the other categorial variable for giving birth in specific age brackets, there are unfortunately not enough instruments for the high number of endogenous regressors. These results can also be seen in table 3.10. The age pattern is much more pronounced now in comparison to the OLS estimates. According to these results, teenage childbearing reduces the probability of finishing high school by 47.3% if the child is conceived before the 15th birthday of the mother. On the other hand, the probability is reduced by only 14.7% for teenage mothers who have conceived before their 18th birthday. Notice that the latter category includes the group of women who have conceived very early. From this it follows that a late teenage childbirth (say at age 18) has a considerably less adverse effect on completing high school than a very early teenage childbirth. This analysis suggests that the model with a binary indicator for a teenage childbirth is misspecified and that one should include the age at birth in the analysis. This result also has important policy

[8]The definition of teenage childbearing still requires that this child is conceived before the 18th birthday.

implications. It suggests that if one is concerned about negative consequences of adolescent childbearing, one should pay special attention to younger teenagers.

I also tried to use miscarriage as an instrument and found that it produced implausible coefficient estimates for the effect of teenage childbearing. The coefficient on teenage childbirth with conception before the age 15 takes on the value of -2.91 predicting that all women with such a young teenage childbirth will successfully finish high school. This is rather implausible. The coefficient on teenage childbirth with conception before the 16th birthday takes on the value -1.20, and the coefficient on teenage childbirth with conception before the 18th birthday is -0.42. These results would suggest that the younger the teenager at the birth of her first child, the more likely she is to graduate from high school, which is implausible. This spurious result can be explained by the following: Of the 703 women who have experienced a miscarriage during their first pregnancy, only 5 have had a teenage childbirth, where the conception was before the 15th birthday. In the first-stage regression, therefore, all women who have experienced a miscarriage will have very low predicted probabilities of teenage childbirth with a conception before the age 15. However, these women who have become pregnant at this very young age may also be at a higher risk of dropping out of high school for other reasons, yet by virtue of the instrument, they will have the lowest predicted probabilities of a teenage childbirth in the sample. Women who have not experienced a miscarriage, on the other hand, have comparably higher predicted probabilities of having a teenage childbirth, yet they are not at an elevated risk of dropping out of high school. Since these predicted probabilities are plugged into the second-stage regression, one is likely to find a relationship between a low predicted probability of a teenage childbirth and having a high chance of not finishing high school. On the other hand, out of the 703 women who have experienced a miscarriage in the first pregnancy around 50 have also had a lifebirth with the age at conception before the 18th birthday in a second or third pregnancy. The relationship between a low predicted probability of not finishing high-school is therefore not as pronounced.

This explanation relies on the existence of some factors that make teenagers who have experienced a miscarriage at a very early age more likely to drop out of high school, and that one has failed to condition on these additional factors. A miscarriage could, for example, proxy for the endogenous decision to initiate sex and become pregnant at a very young age which is more likely to result in miscarriages as shown before. This decision may be correlated with low educational achievement due to some other unobserved factors. This would invalidate miscarriage as an instrument. In a 2SLS regression, one should therefore condition on age at first conception. Otherwise it is an omitted variable correlated with the instrument miscarriage leading to biased estimates. In table 3.11, I present 2SLS estimates where I treat a young conception as an additional variable. Young conception takes the value 1 if the conception was before the mother's 15th birthday, 0 otherwise. In the first column, a young age at conception is treated as exogenous, and I use miscarriage as instrument for teenage childbirth. The coefficient to teenage childbirth raises by about 0.1. At the same time, a young age at conception is associated with a highly increased risk of dropping out of high school. In the second column, I use miscarriage and age at menarche as instruments while still treating young age at conception as exogenous. The coefficient estimates are qualitatively comparable to the result in the first column. With two instruments, one can also test the overidentifying restrictions. Without including young age at conception, they have been earlier rejected when using miscarriage and age at menarche as instruments. Now after conditioning on age at conception, one cannot reject these overidentifying restrictions. Conditioning on a young age at conception renders miscarriage and age at menarche valid instruments judged by Hansen's J statistic. Again, young age at conception is associated with a big increase in the risk of dropping out of high school. In the third column, I treat both young age at conception and teenage childbirth as endogenous variables. The coefficient estimates suggest that only very young childbearing has negative consequences for high school completion whereas a teenage childbirth itself may even be beneficial. For all three specifications the F-tests of the instruments are very high in the first-stage regression suggesting that conditioning on young age at conception also mitigates any possible weak instrument problems.

3.2 Conclusion

The goal of this essay was to explore in an empirical application why coefficient estimates using alternative sets of instruments differ. I focused on three possible explanations.

The first explanation are defective instruments. If one instrument is valid and relevant and the other instrument is not, it is not surprising that they result in differing coefficient estimates on the endogenous regressor. In this essay, I find inconclusive evidence on whether the instruments are defective. While there is no sign that instrument weakness is a problem, there are some signs that the instruments cannot be safely excluded from the main equation, especially for miscarriage. Using Sargan's test and Hansen's J statistic, I can reject the overidentifying restrictions when I use both the occurrence of a miscarriage and age at menarche (or a dummy for an early menarche) as instruments for the whole sample. Testing the set of instruments age at menarche, its square, and miscarriage, however, I cannot reject the overidentifying restrictions. In addition, if I only use age at menarche and its square as instruments, I again cannot reject the overidentifying restrictions. Hotz et al. (1997, 1999) restrict their sample to women who have become pregnant as teenagers. Using this sample definition, I cannot reject the overidentifying restrictions. Because in this sample all women have been pregnant, there is no correlation between pregnancies and the occurrence of a miscarriage which may have been the problem in the whole sample. In the subsample of ever pregnant women and women who have children, one cannot reject the overidentifying restrictions, probably for the same reasons.

The second explanation relies on treatment effect heterogeneity where different instruments identify the treatment effect parameters for different subpopulations as defined by their probability of experiencing an early childbirth. If there is treatment effect heterogeneity in this sense, the function linking the marginal treatment effect and the participation probability is not constant and the alternative instruments sweep out different parts of this distribution. Indeed, if one compares the effects of the two instruments on the probability of having a teenage childbirth, one finds that the miscarriage has a quantitatively stronger effect and therefore the predicted probabilities using age at menarche and miscarriage only partly overlap. However, using Moffitt's model, I do not find indications for important treatment effect heterogeneity. The results from these model estimates are consistent with homogenous treatment effects. With constant treatment effects all valid and relevant instruments should identify the same parameter. Thus, I conclude that this kind of treatment effect heterogeneity is unlikely to explain the differences in the 2SLS coefficient estimates when using alternative sets of instruments.

A third explanation for the differences in instrumental variable estimates focuses on the exact mechanism by which the instruments affects the outcome (Moffitt 2005). In this particular application both a miscarriage and a high age at menarche make a postponement of first birth after the age of 18 more likely. A miscarriage affects teenagers who have already been pregnant while a high age at menarche shortens the time at risk for teenagers to become pregnant. Implicit in the use of a binary indicator for a teenage childbirth is the assumption that the effect of a teenage childbirth does not depend on the exact age the teenager gives birth to the child. However, this assumption is open to testing, and I find that the adverse effect of teenage childbearing on educational outcomes seems to be stronger for very young teenagers. Since miscarriage disproportionately affects this age group, this could be an additional explanation for the differences in coefficient estimates when using alternative sets of instruments. If these age effects are an important factor, and the instruments have asymmetric effects on subpopulations as defined by their age at first birth, then the model with only a binary indicator for teenage childbirth is misspecified.

A miscarriage may also directly alter attitudes towards sexual activity and contraception and lead to different expectations about the timing of future births which itself also may directly affect educational outcomes. Furthermore, all these instruments may plausibly also influence marital prospects if, for example, the father would have felt compelled to marry the woman had no miscarriage occurred. Marital prospects themselves can change the incentives for educational achievement. If there are multiple potential mechanisms, and one only includes one mechanism in the outcome equation, then the instruments may potential serve as proxies for the omitted mechanisms resulting in a misspecification of the model.

This essay explored empirically reasons for why different instrumental variables estimates differ. While I do not find indications that differences in the marginal treatment effects are the main cause of differences in these estimates, I find evidence that there may be some problems with the validity of the instruments, especially with miscarriage.

Furthermore, I find that a binary indicator for teenage childbearing masks important differences in the effect for different age groups. The adverse effect on educational achievement is much stronger for very young mothers in comparison to mothers who gave birth at age 17 or 18. Part of the differences in the estimates may be explained by the fact that having a miscarriage disproportionately affects very young mothers, and this may also be the reason why miscarriage is not a valid instrument for a binary indicator for teenage childbearing if one does not also conditions on young age at conception. These results also have important policy implications. If one is concerned about adverse consequences of adolescent childbearing on educational outcomes for mothers, one should focus the attention on the very young age group since here the adverse effects are strongest. The message for the practitioner is that if one encounters differences in estimates using alternative sets of instruments, it is fruitful to look for reasons for these differences. These reasons may be defects of instruments, treatment effect heterogeneity, or as it was the case here to extend the analysis and look more closely at observable differences in subpopulations affected by the instruments.

Table 3.1: Summary Statistics. Means of Variables

	No teenage birth ($D = 0$)	Teenage birth ($D = 1$)
Less than high school	0.14	0.51
Age at Interview	31.75	31.39
Race[a]		
Black	0.19	0.37
White	0.72	0.53
Religion[b]		
No religion	0.15	0.16
Catholic	0.29	0.29
Protestant	0.49	0.52
Family Background		
Born in USA	0.82	0.82
Not Intact Family	0.33	0.51
No father present	0.06	0.12
Mother's Education[c]		
No HS	0.25	0.42
High school	0.34	0.33
Some college	0.22	0.13
BA+	0.18	0.06
Father's Education[d]		
No HS	0.23	0.35
High school	0.28	0.28
Some college	0.17	0.11
BA+	0.24	0.07
Instruments		
Early Menarche	0.49	0.60
Age at menarche	12.63	12.18
Age at menarche squared	162.36	151.20
Miscarriage	0.12	0.06
Number of Observations	5534	909

a: Omitted Category is other race.

b: Omitted Category is other religion.

c: Omitted Category is no mother present, refusal of answer, or individual does not know.

d: Omitted Category is refusal of answer, or individual does not know.

Table 3.2: OLS Regressions of the Instruments on other Covariates

	Miscarriage	Miscarriage for ever pregnant women	Miscarriage	Early menarche	Age at menarche
Pregnant before 15			0.057		
			(0.023)		
Age at interv.	0.003	0.001	0.003	-0.004	0.015
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
Race[a]					
Black	-0.010	-0.016	-0.012	0.028	-0.078
	(0.017)	(0.021)	(0.017)	(0.027)	(0.089)
White	0.007	0.018	0.008	0.001	0.042
	(0.015)	(0.019)	(0.015)	(0.023)	(0.079)
Religion[b]					
No religion	0.010	0.021	0.010	0.025	-0.054
	(0.019)	(0.027)	(0.019)	(0.031)	(0.104)
Catholic	-0.005	-0.011	-0.004	0.008	0.021
	(0.018)	(0.024)	(0.018)	(0.029)	(0.096)
Protestant	0.014	0.013	0.014	0.026	-0.041
	(0.018)	(0.024)	(0.018)	(0.028)	(0.094)
Family Background					
Born in USA	0.002	0.005	0.001	0.052	-0.237
	(0.011)	(0.015)	(0.011)	(0.018)	(0.061)
No int. family	0.022	0.012	0.021	-0.006	-0.079
	(0.009)	(0.012)	(0.009)	(0.014)	(0.048)
No father pres.	-0.003	0.007	-0.002	0.006	0.232
	(0.028)	(0.035)	(0.023)	(0.045)	(0.150)

Table continued on following page

Table 3.2 continued

	Mis-carriage	Miscarriage for ever pregnant women	Mis-carriage	Early Menarche	Menarche
Mother's educ.[c]					
No HS	0.017	0.014	0.019	0.017	-0.302
	(0.030)	(0.037)	(0.030)	(0.047)	(0.159)
High school	0.020	0.026	0.021	-0.003	-0.021
	(0.029)	(0.037)	(0.029)	(0.047)	(0.158)
Some college	0.013	0.027	0.015	0.013	-0.300
	(0.030)	(0.038)	(0.030)	(0.048)	(0.161)
BA+	-0.003	0.012	-0.001	-0.010	-0.311
	(0.031)	(0.039)	(0.031)	(0.049)	(0.165)
Father's educ.[d]					
No HS	0.011	0.015	0.013	-0.023	0.236
	(0.025)	(0.031)	(0.025)	(0.040)	(0.134)
High school	-0.001	-0.003	0.001	-0.032	0.264
	(0.025)	(0.031)	(0.025)	(0.040)	(0.133)
Some college	0.005	0.015	0.008	-0.050	0.349
	(0.026)	(0.032)	(0.026)	(0.041)	(0.138)
BA+	0.003	0.024	0.007	-0.070	0.447
	(0.028)	(0.032)	(0.026)	(0.041)	(0.139)
Constant	-0.028	0.072	-0.035	0.606	12.276
	(0.043)	(0.056)	(0.023)	(0.069)	(0.232)
Number of obs.	6443	4810	6443	6443	6443

Notes: Standard errors in parentheses.

a: Omitted Category is other race.

b: Omitted Category is other religion.

c: Omitted Category is no mother present, refusal of answer, or individual does not know.

d: Omitted Category is refusal of answer, or individual does not know.

Table 3.3: OLS Regression. Determinants of Teenage Childbirth (Dummy)

	I	II	III	IV	V	VI	VII
Instruments							
Miscarriage	-0.082			-0.084	-0.085		-0.085
	(0.013)			(0.013)	(0.013)		(0.013)
Early men.		0.044		0.045			
		(0.008)		(0.008)			
Menarche			-0.016		-0.016	-0.010	-0.014
			(0.002)		(0.002)	(0.023)	(0.023)
Menarche^2						-0.0002	-0.0001
						(0.001)	(0.001)
Controls							
Age at interv.	-0.002	-0.002	-0.002	-0.001	-0.001	-0.002	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Race[a]							
Black	0.053	0.052	0.052	0.051	0.051	0.052	0.051
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
White	-0.048	-0.049	-0.048	-0.048	-0.047	-0.048	-0.047
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Religion[b]							
No religion	0.029	0.027	0.027	0.028	0.028	0.027	0.028
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Catholic	0.028	0.028	0.029	0.028	0.029	0.029	0.029
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Protestant	0.019	0.017	0.018	0.018	0.019	0.018	0.019
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Family Background							
Born in USA	0.046	0.043	0.042	0.043	0.042	0.042	0.042
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
No int. family	0.064	0.063	0.061	0.064	0.063	0.061	0.063
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
No father pres.	-0.097	-0.097	-0.093	-0.097	-0.093	-0.093	-0.093
	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)

Table continued on following page

Table 3.3 cont'd

	I	II	III	IV	V	VI	VII
Mother's educ.[c]							
No HS	-0.105	-0.107	-0.111	-0.105	-0.101	-0.111	-0.110
	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
High school	-0.161	-0.163	-0.166	-0.161	-0.165	-0.166	-0.165
	(0.032)	(0.032)	(0.032)	(0.031)	(0.031)	(0.032)	(0.031)
Some College	-0.199	-0.201	-0.205	-0.200	-0.204	-0.205	-0.204
	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
BA+	-0.207	-0.207	-0.212	-0.207	-0.212	-0.212	-0.212
	(0.032)	(0.033)	(0.033)	(0.032)	(0.032)	(0.033)	(0.033)
Father's educ.[d]							
No HS	-0.075	-0.075	-0.072	-0.073	-0.071	-0.072	-0.071
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
High school	-0.118	-0.116	-0.114	-0.116	-0.114	-0.114	-0.114
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Some College	-0.147	-0.146	-0.142	-0.145	-0.142	-0.142	-0.142
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
BA+	-0.170	-0.167	-0.163	-0.167	-0.163	-0.163	-0.163
	(0.028)	(0.030)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
Constant	0.419	0.395	0.614	0.392	0.616	0.580	0.601
	(0.046)	(0.047)	(0.056)	(0.047)	(0.055)	(0.153)	(0.152)
Number of obs.	6443	6443	6443	6443	6443	6443	6443

Notes: Standard errors in parentheses.

a: Omitted Category is other race.

b: Omitted Category is other religion.

c: Omitted Category is no mother present, refusal of answer, or individual does not know.

d: Omitted Category is refusal of answer, or individual does not know.

Table 3.4: Probit Regression. Determinants of Teenage Childbirth (Dummy)

	I	II	III	IV	V	VI	VII
Instruments							
Miscarriage	-0.497			-0.506	-0.508		
	(0.079)			(0.079)	(0.079)		
Early menarche		0.227		0.233			
		(0.042)		(0.042)			
Menarche			-0.080		-0.082	0.069	0.051
			(0.012)		(0.012)	(0.121)	(0.121)
Menarche^2						-0.006	-0.005
						(0.005)	(0.005)
Controls							
Age at interv.	-0.009	-0.008	-0.008	-0.008	-0.007	-0.008	-0.007
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Race[a]							
Black	0.210	0.208	0.210	0.207	0.207	0.212	0.209
	(0.082)	(0.082)	(0.082)	(0.082)	(0.082)	(0.082)	(0.082)
White	-0.221	-0.229	-0.227	-0.225	-0.223	-0.229	-0.225
	(0.075)	(0.075)	(0.075)	(0.075)	(0.075)	(0.075)	(0.075)
Religion[b]							
No religion	0.191	0.173	0.179	0.187	0.192	0.178	0.191
	(0.115)	(0.116)	(0.116)	(0.116)	(0.117)	(0.116)	(0.117)
Catholic	0.200	0.199	0.204	0.201	0.206	0.204	0.206
	(0.109)	(0.109)	(0.110)	(0.110)	(0.110)	(0.110)	(0.110)
Protestant	0.148	0.137	0.141	0.145	0.149	0.142	0.150
	(0.107)	(0.107)	(0.108)	(0.108)	(0.108)	(0.108)	(0.108)
Family Background							
Born in USA	0.192	0.172	0.167	0.176	0.170	0.167	0.170
	(0.060)	(0.060)	(0.060)	(0.061)	(0.061)	(0.060)	(0.061)
No int. family	0.323	0.312	0.306	0.328	0.321	0.307	0.322
	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)
No father pres.	-0.356	-0.355	-0.339	-0.359	-0.343	-0.339	-0.342
	(0.123)	(0.122)	(0.122)	(0.123)	(0.123)	(0.122)	(0.123)

Table continued on following page

Table 3.4 cont'd

	I	II	III	IV	V	VI	VII
Mother's ed.[c]							
No HS	-0.312	-0.326	-0.346	-0.321	-0.340	-0.347	-0.341
	(0.126)	(0.126)	(0.126)	(0.126)	(0.127)	(0.126)	(0.127)
High school	-0.537	-0.543	-0.561	-0.539	-0.557	-0.564	-0.560
	(0.126)	(0.126)	(0.126)	(0.126)	(0.126)	(0.126)	(0.126)
Some College	-0.742	-0.749	-0.771	-0.748	-0.770	-0.773	-0.773
	(0.131)	(0.131)	(0.132)	(0.132)	(0.132)	(0.132)	(0.132)
BA+	-0.839	-0.834	-0.862	-0.844	-0.872	-0.862	-0.872
	(0.140)	(0.140)	(0.140)	(0.140)	(0.141)	(0.140)	(0.141)
Father's ed.[d]							
No HS	-0.206	-0.199	-0.186	-0.199	-0.186	-0.189	-0.189
	(0.108)	(0.108)	(0.108)	(0.108)	(0.108)	(0.108)	(0.108)
High school	-0.397	-0.386	-0.373	-0.391	-0.378	-0.376	-0.380
	(0.109)	(0.108)	(0.108)	(0.109)	(0.109)	(0.108)	(0.109)
Some College	-0.564	-0.556	-0.540	0.559	-0.542	-0.542	-0.544
	(0.117)	(0.118)	(0.117)	(0.117)	(0.117)	(0.117)	(0.117)
BA+	-0.821	-0.801	-0.783	-0.810	-0.791	-0.787	-0.795
	(0.123)	(0.122)	(0.122)	(0.123)	(0.123)	(0.122)	(0.123)
Constant	-0.191	-0.334	0.775	-0.340	0.800	-0.134	-0.013
	(0.209)	(0.211)	(0.258)	(0.211)	(0.259)	(0.779)	(0.781)
Log-L	-2338.6	-2345.8	-2339.7	-2322.9	-2316.7	-2338.9	-2316.0
Number of obs.	6443	6443	6443	6443	6443	6443	6443

Notes: Standard errors in parentheses.

a: Omitted Category is other race.

b: Omitted Category is other religion.

c: Omitted Category is no mother present, refusal of answer, or individual does not know.

d: Omitted Category is refusal of answer, or individual does not know.

Table 3.5: 2SLS Regressions. Dependent Variable No High School

Set of instruments	Coeff. on teenage childb.	S.E.	F-stat	Sargan-stat/ Hansen's J
OLS	0.291***	0.013		
Miscarriage	-0.418***	0.206	37.96***[a]	
Age at menarche	0.147	0.168	39.53***[a]	
Early menarche	0.271	0.200	27.50***[a]	
Miscarriage, Age at menarche	-0.130	0.127	39.96***[b]	4.729**[c] /4.352**[c]
Miscarriage, Early menarche	-0.127	0.138	33.67***[b]	5.811**[c] /5.662**[c]
Age at menarche, age at menarche^2	0.145	0.168	19.79***[b]	0.082[c]/0.086[c]
Miscarriage, age at menarche, age at menarche^2	-0.122	0.132	33.44***[d]	4.759*[e]/4.352[e]

All 2SLS regressions also include controls for age at interview, race, religion, dummies for intact family background, parental education, and whether the woman was born in the United States.

a: F-statistic with F(1, 6424): Test of significance of instrument in first stage regression.

b: F-statistic with F(2, 6423): Test of joint significance of instruments in first stage regression.

c: Chi-squared with 1 degree of freedom: Test of overidentifying restrictions.

d: F-statistic with F(3, 6422): Test of joint significance of instruments in first stage regression.

e: Chi-squared with two degrees of freedom: Test of overidentifying restrictions

*** significant at 1% ** significant at 5% *significant at 10%

Table 3.6: 2SLS Regressions on Different Subsamples. Dependent Variable No High School

Set of instruments	Coeff. on teenage childb.	S.E.	F-stat	Sargan-stat/ Hansen's J
Teenage Pregnancy Sample				
OLS	0.177***	0.029		
Miscarriage	-0.123	0.083	195.70***[a]	
Age at menarche	0.183	0.608	2.89*[a]	
Early menarche	0.601	0.526	4.51**[a]	
Miscarriage, Age at menarche	-0.116	0.081	101.60***[b]	0.229[c] /0.219[c]
Ever Pregnant Sample				
OLS	0.267***	0.015		
Miscarriage	-0.044	0.125	72.01***[d]	
Age at menarche	0.107	0.154	43.80***[d]	
Early menarche	0.274	0.177	32.29***[d]	
Miscarriage, Age at menarche	0.013	0.096	59.97***[e]	0.562[c] /0.542[c]
Have Kids Sample				
OLS	0.260***	0.015		
Miscarriage	-0.160	0.166	42.03***[f]	
Age at menarche	0.040	0.151	44.84***[f]	
Early menarche	0.240	0.164	36.18***[f]	
Miscarriage, Age at menarche	-0.057	0.110	44.79***[g]	0.793[c] /0.768[c]

All 2SLS regressions also include controls for age at interview, race, religion, dummies for intact family background, parental education, and whether the woman was born in the United States.

a: F-statistic with F(1, 1265): Test of significance of instrument.

b: F-statistic with F(2, 1264): Test of joint significance of instruments.

c: Chi-squared with 1 degree of freedom: Test of overidentifying restrictions.

d: F-statistic with F(1, 4791): Test of significance of instrument.

e: F-statistic with F(2, 4792): Test of joint significance of instruments.

f: F-statistic with F(1, 4254): Test of significance of instrument.

g: F-statistic with F(2, 4253): Test of joint significance of instruments.

*** significant at 1% ** significant at 5% *significant at 10%

Table 3.7: Effect of Teenage Childb. on HS Completion in Hotz et al. and Ribar

Set of Instruments	Hotz et al. (1999)				Ribar (1994)
	High School Diploma		High School Diploma or GED		High School Diploma or GED
OLS	0.41[a]	(0.03)[a]	0.19[a]	(0.03)[a]	0.234[c]
Hotz et al.					
Miscarriage	0.16[b]	(0.09)[b]	-0.03[b]	(0.08)[b]	-
Ribar					
Age at menarche	-	-	-	-	0.711[d]
Ob-Gyn availability	-	-	-	-	0.398[d]
Abortion rate	-	-	-	-	negative but small

Notes: Standard errors for Hotz et al. (1999) in parentheses.

a: Based on Hotz et al. (1999), table 9.

b: Based on Hotz et al. (1999), table 4. Results are for teenage pregnancy sample.

c: Based on Ribar (1994). Teenage childbearing exogenous in probit model.

d: Based on Ribar (1994), Footnote 18.

Table 3.8: Moffitt's Model without Heterogeneity. Dep. Variable: Less than HS

	I	II	III	IV
γ_0	0.245	-0.117	0.202	-0.120
	(0.157)	(0.118)	(0.151)	(0.116)
γ_1				
Age at interv.	-0.003	-0.003	-0.003	-0.003
	(0.001)	(0.001)	(0.001)	(0.001)
Race[a]				
Black	-0.014	0.002	-0.013	0.002
	(0.022)	(0.022)	(0.022)	(0.022)
White	0.018	0.001	0.016	0.001
	(0.019)	(0.020)	(0.019)	(0.020)
Religion[b]				
No religion	0.064	0.069	0.065	0.069
	(0.020)	(0.021)	(0.020)	(0.021)
Catholic	0.079	0.084	0.079	0.084
	(0.018)	(0.019)	(0.018)	(0.019)
Protestant	0.042	0.044	0.042	0.044
	(0.017)	(0.018)	(0.017)	(0.018)
Family Background				
Born in USA	-0.093	-0.073	-0.091	-0.073
	(0.015)	(0.016)	(0.015)	(0.016)
No int. family	0.034	0.054	0.037	0.054
	(0.013)	(0.012)	(0.013)	(0.012)
No father pres.	-0.082	-0.116	-0.086	-0.116
	(0.042)	(0.045)	(0.042)	(0.045)
Mother's educ.[c]				
No HS	-0.115	-0.152	-0.120	-0.153
	(0.048)	(0.050)	(0.048)	(0.050)
High school	-0.245	-0.303	-0.252	-0.304
	(0.051)	(0.051)	(0.051)	(0.051)
Some college	-0.290	-0.361	-0.299	-0.361
	(0.055)	(0.053)	(0.055)	(0.053)
BA+	-0.296	-0.371	-0.305	-0.372
	(0.056)	(0.054)	(0.055)	(0.054)

Table cont'd on following page

Table 3.8 cont'd

	I	II	III	IV
Father's educ.[d]				
No HS	-0.090	-0.118	-0.093	-0.118
	(0.038)	(0.041)	(0.038)	(0.041)
High school	-0.166	-0.209	-0.171	-0.209
	(0.039)	(0.041)	(0.039)	(0.041)
Some college	-0.212	-0.263	-0.218	-0.263
	(0.042)	(0.043)	(0.041)	(0.042)
BA+	-0.234	-0.293	-0.241	-0.294
	(0.043)	(0.043)	(0.043)	(0.043)
Constant	0.625	0.780	0.645	0.783
	(0.088)	(0.080)	(0.086)	(0.080)
Number of obs.	6443	6443	6443	6443
Instruments	Age at men.	Miscarriage, age at men.	Age at men., age at men.^{2}	Age at men., age at men.^{2}, miscarriage

Notes: Standard error in parentheses

a: Omitted Category is other race.

b: Omitted Category is other religion.

c: Omitted Category is no mother present, refusal of answer, or individual does not know.

d: Omitted Category is refusal of answer, or individual does not know

Table 3.9: Moffitt's Model with Heterogeneity. Dep. Variable: Less than HS

	I	II	III	IV
γ_0	0.599	-0.032	0.529	0.041
	(0.340)	(0.221)	(0.309)	(0.211)
γ_1	-0.641	-0.193	-0.608	-0.367
	(0.558)	(0.424)	(0.528)	(0.415)
Age at interv.	-0.002	-0.003	-0.003	-0.003
	(0.001)	(0.001)	(0.001)	(0.081)
Race[a]				
Black	-0.018	0.002	-0.015	0.002
	(0.023)	(0.022)	(0.022)	(0.022)
White	0.023	0.002	0.021	0.003
	(0.020)	(0.020)	(0.020)	(0.020)
Religion[b]				
No religion	0.062	0.067	0.063	0.067
	(0.020)	(0.021)	(0.020)	(0.021)
Catholic	0.076	0.083	0.078	0.082
	(0.018)	(0.019)	(0.018)	(0.019)
Protestant	0.041	0.043	0.042	0.042
	(0.018)	(0.018)	(0.017)	(0.018)
Family Background				
Born in USA	-0.097	-0.073	-0.094	-0.072
	(0.016)	(0.016)	(0.015)	(0.016)
No int. fam.	0.027	0.053	0.030	0.052
	(0.014)	(0.013)	(0.014)	(0.012)
No father pres.	-0.083	-0.118	-0.089	-0.120
	(0.042)	(0.046)	(0.043)	(0.046)
Mother's educ.[c]				
No HS	-0.121	-0.156	-0.130	-0.161
	(0.049)	(0.051)	(0.050)	(0.052)
High school	-0.248	-0.307	-0.258	-0.312
	(0.052)	(0.053)	(0.052)	(0.053)
Some college	-0.285	-0.365	-0.298	-0.369
	(0.055)	(0.054)	(0.055)	(0.055)
BA+	-0.288	-0.373	-0.301	-0.377
	(0.056)	(0.054)	(0.056)	(0.055)

Table cont'd on following

Table 3.9 cont'd

	I	II	III	IV
Father's educ.[d]				
No HS	-0.094	-0.121	-0.099	-0.123
	(0.038)	(0.041)	(0.039)	(0.042)
High school	-0.166	-0.211	-0.173	-0.214
	(0.040)	(0.042)	(0.040)	(0.042)
Some college	-0.207	-0.264	-0.216	-0.266
	(0.040)	(0.043)	(0.042)	(0.043)
BA+	-0.220	-0.293	-0.231	-0.294
	(0.046)	(0.044)	(0.042)	(0.044)
Constant	0.589	0.779	0.615	0.781
	(0.096)	(0.081)	(0.093)	(0.081)
Number of obs.	6443	6443	6443	6443
Instruments	Age at men.	Miscarriage, age at men.	Age at men., age at men.^{2}	Miscarriage, age at men., age at men.^{2}

Notes: Standard errors in parentheses

a: Omitted Category is other race.

b: Omitted Category is other religion.

c: Omitted Category is no mother present, refusal of answer, or individual does not know.

d: Omitted Category is refusal of answer, or individual does not know

Table 3.10: OLS and 2SLS Regressions. Different Definitions of Teenage Childbirth. Dep. Variable Less than High School

	OLS					2SLS			
Conception before[b]									
age 15	0.359					0.473			
	(0.034)					(0.554)			
age 16		0.341					0.203		
		(0.022)					(0.236)		
age 17			0.347					0.152	
			(0.016)					(0.175)	
age 18[c]				0.291					0.147
				(0.013)					(0.356)
Age at first teen childbirth									
before age 15					0.490				
					(0.049)				
age 15-16					0.355				
					(0.035)				
age 16-17					0.376				
					(0.026)				
age 17-18					0.287				
					(0.021)				
older than 18					0.182				
					(0.022)				
Observations	6443	6443	6443	6443	6443	6443	6443	6443	6443

Notes: Standard errors in parentheses.

Age at menarche used as instrument in 2SLS results. All regressions also include controls for age at interview, race, religion, dummies for intact family background, parental education, and whether the woman was born in the United States. Age at first teenage childbirth before age 18 requires conception before the 18th birthday.

Table 3.11: 2SLS Regressions Including Young Age at Conception as Exogenous and Endogenous Regressor

	I	II	III
Endogenous:			
Teenage childbirth	-0.316	-0.248	-0.275
	(0.173)	(0.142)	(0.149)
Young age at conception			0.702
			(0.335)
Exogenous:			
Young age at conception	0.500	0.463	
	(0.098)	(0.087)	
F-test	78.43	46.52	39.96/ 30.55
Hansen's J		0.533	
Instruments used	Miscarriage	Miscarriage	Miscarriage
		Age at menarche	Age at menarche

Notes: All F-tests are F(2, 6423) and test significance of instruments in first-stage regression. F-test in column III are for the first stage regressions for teenage childbirth/ young age at conception respectively. Hansen's J statistic is Chi squared with one degree of freedom.

Figure 3.1: Predicted Probability of a Teenage Childbirth (OLS and Probit) Using Miscarriage and Age at Menarche as Instruments

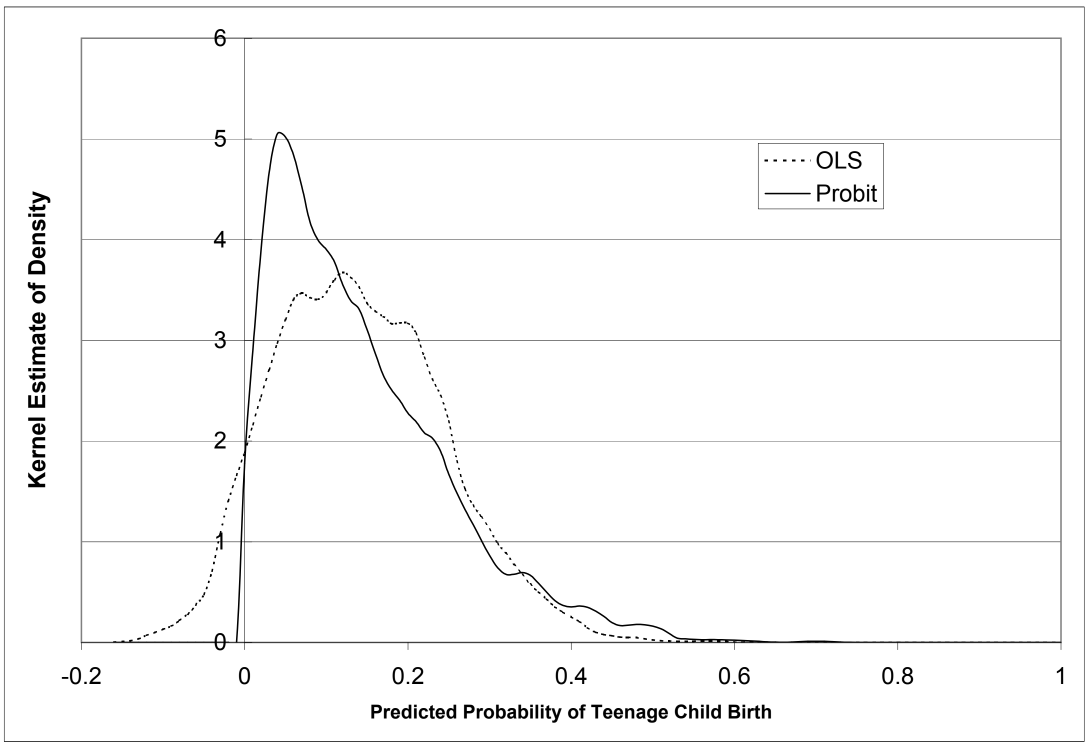


Figure 3.2: Predicted Probabilities of Teenage Childbirth (Probit) by Values of the Instruments (Other Covariates are Held Fixed)

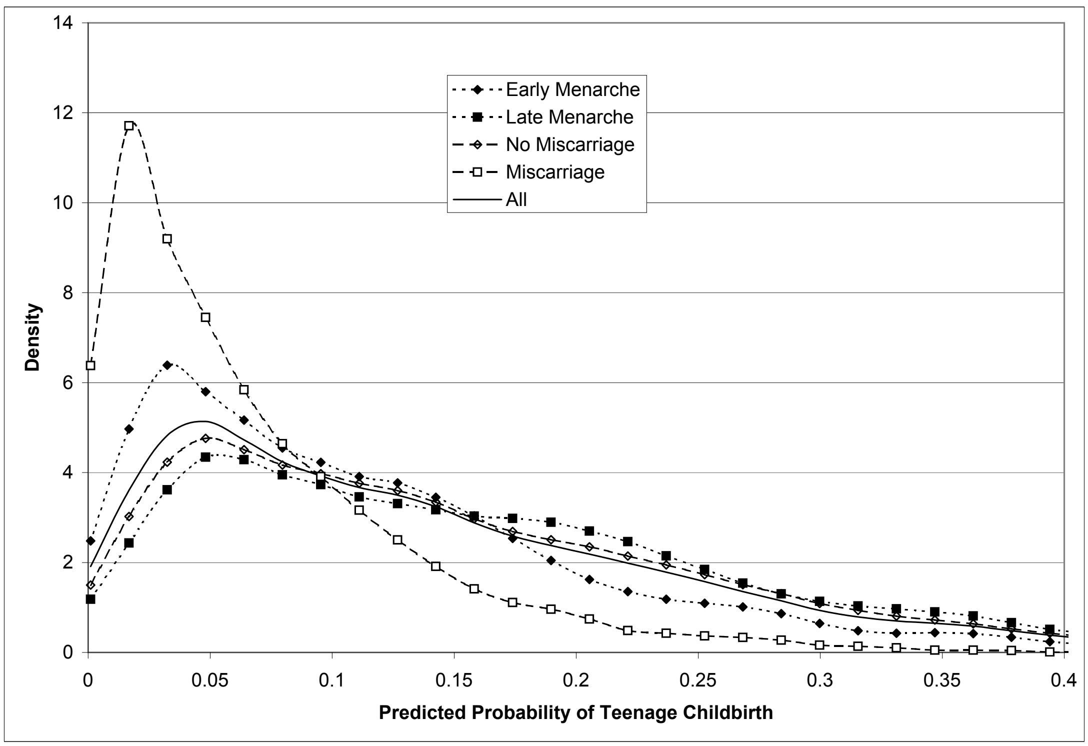


Figure 3.3: Box-Plots at 4 Quartiles of the Predicted Probabilities for Having a Teenage Childbirth Setting Instrument Values to Alternative Values

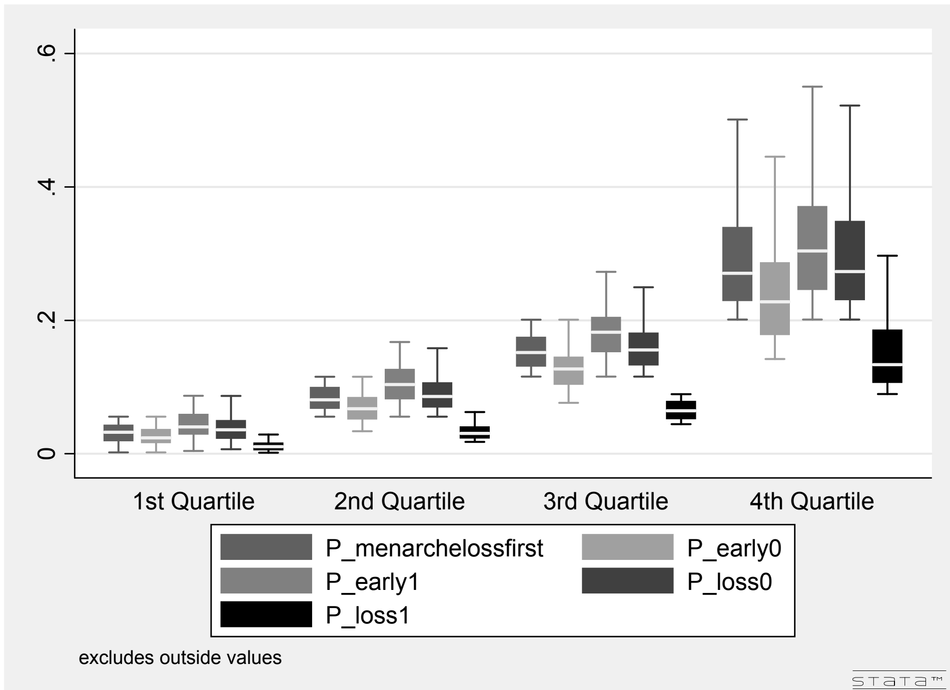


Figure 3.4: MTE Implied by Moffitt's model Using Age at Menarche as an Instrument

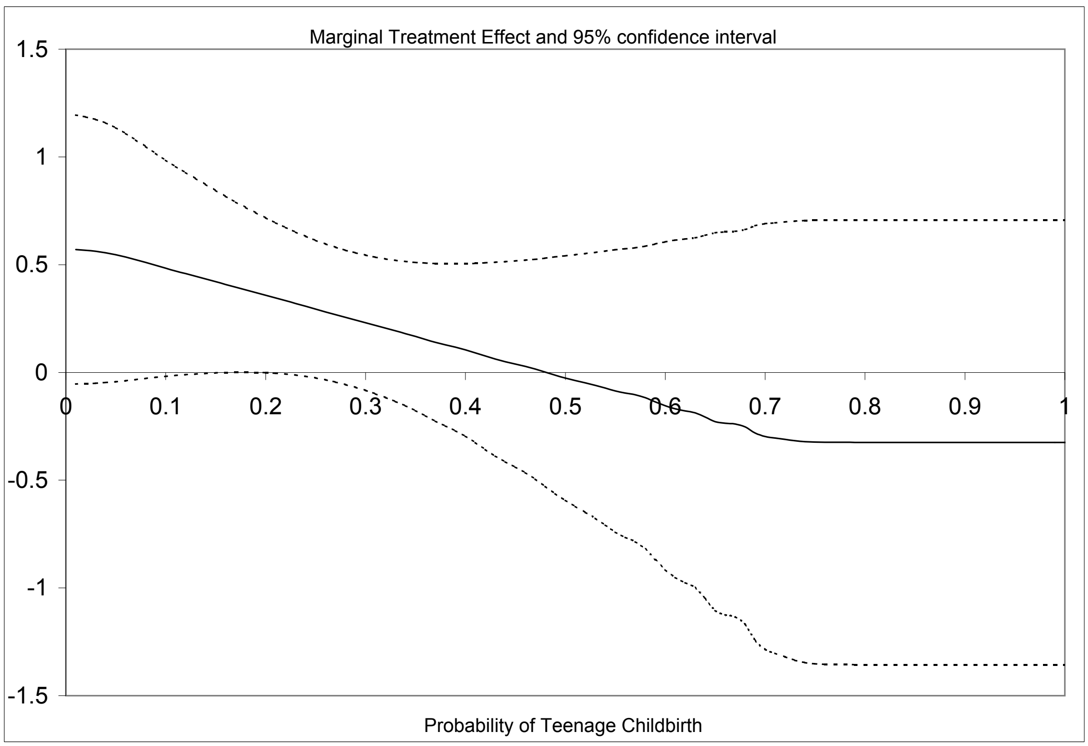


Figure 3.5: MTE Implied by Moffitt's Model Using Age at Menarche and Miscarriage as Instruments

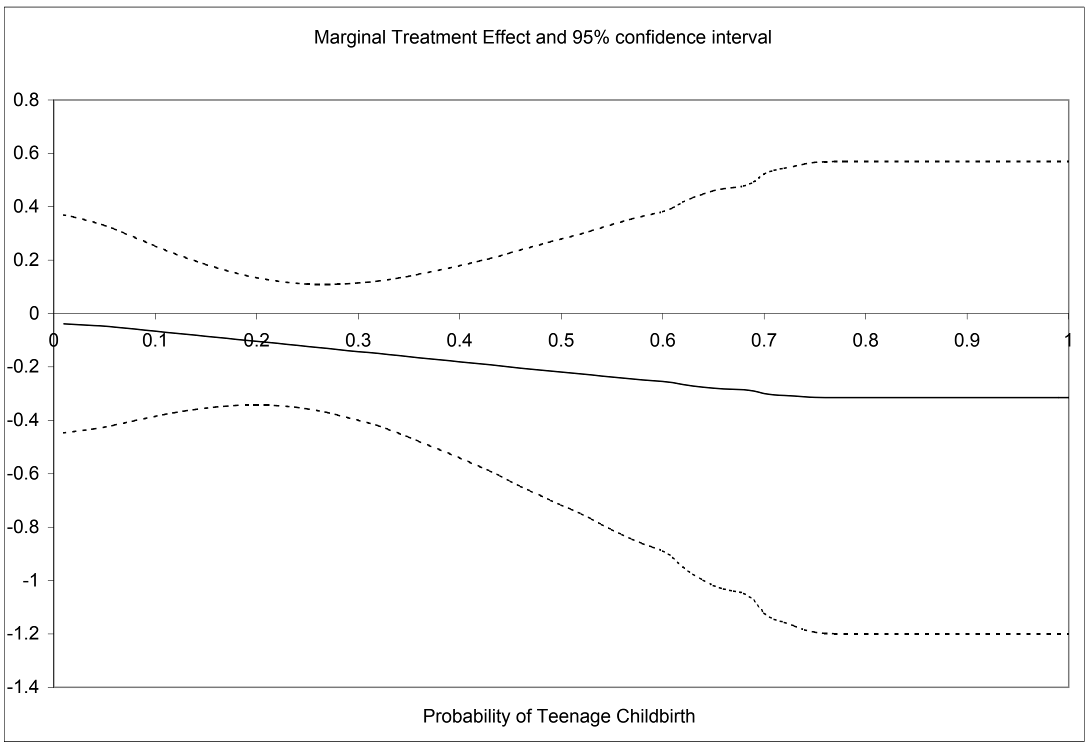


Figure 3.6: MTE Implied by Moffitt's Model Using Age at Menarche and its Square as Instruments

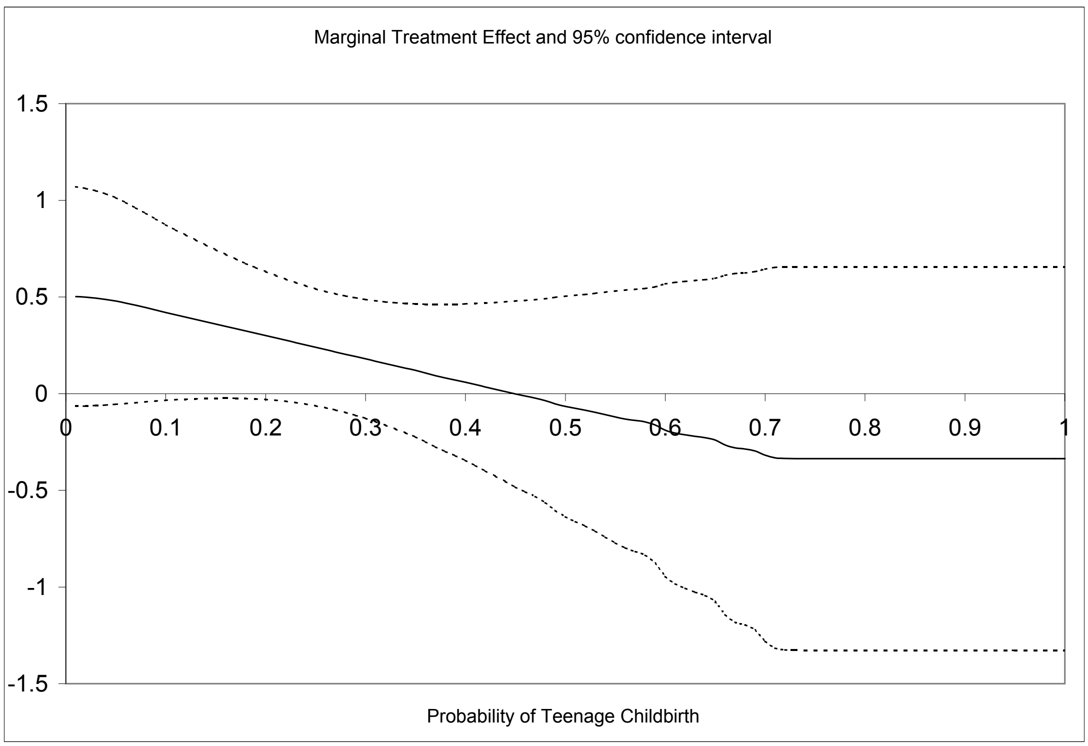


Figure 3.7: MTE implied by Moffitt's model using Age at Menarche, its Square, and Miscarriage as Instruments

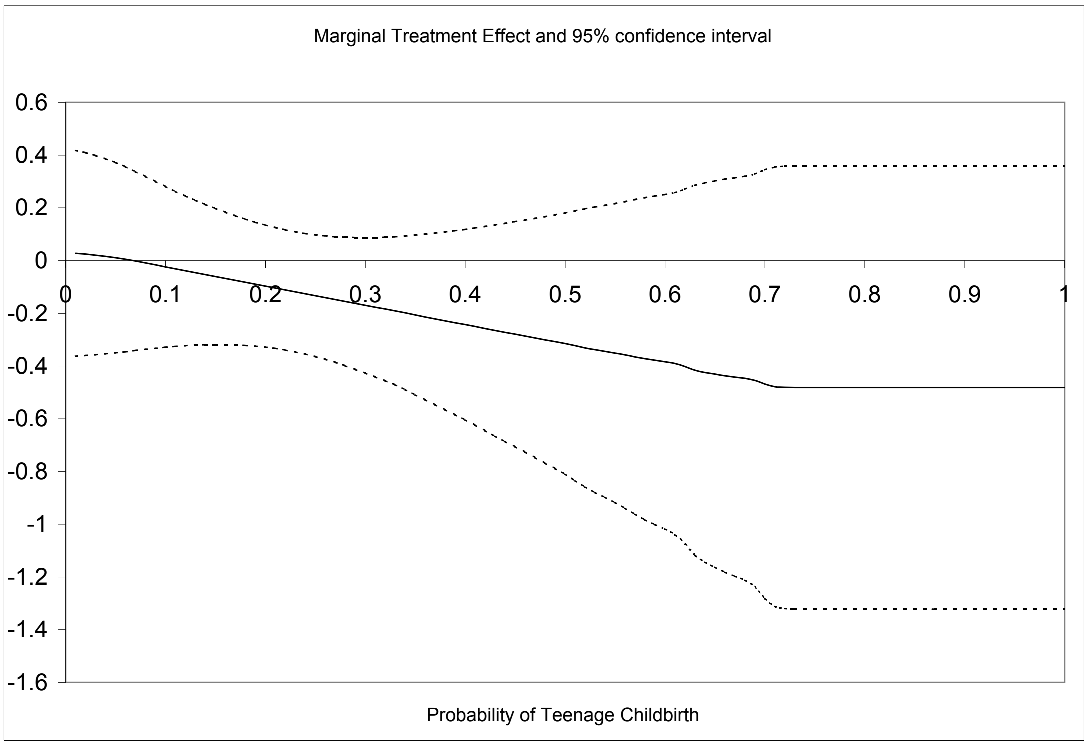


Figure 3.8: Kernel Density Estimates of Age at First Conception for all Women and for Women with Early (< 13 years) and Late Menarche

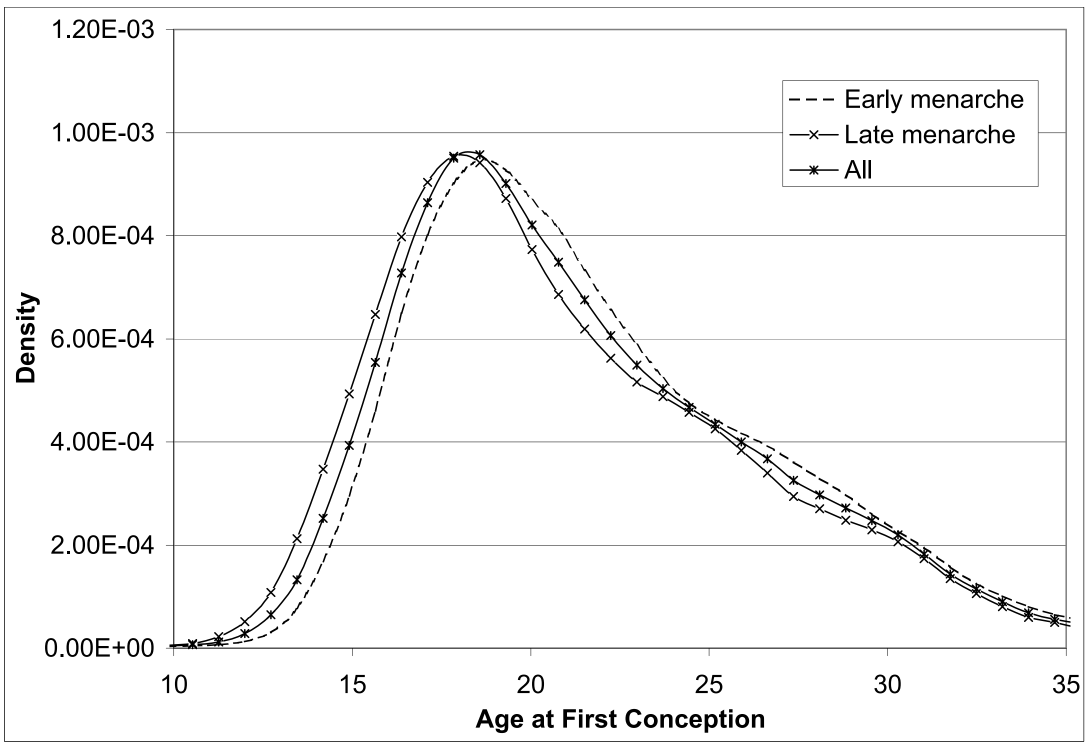
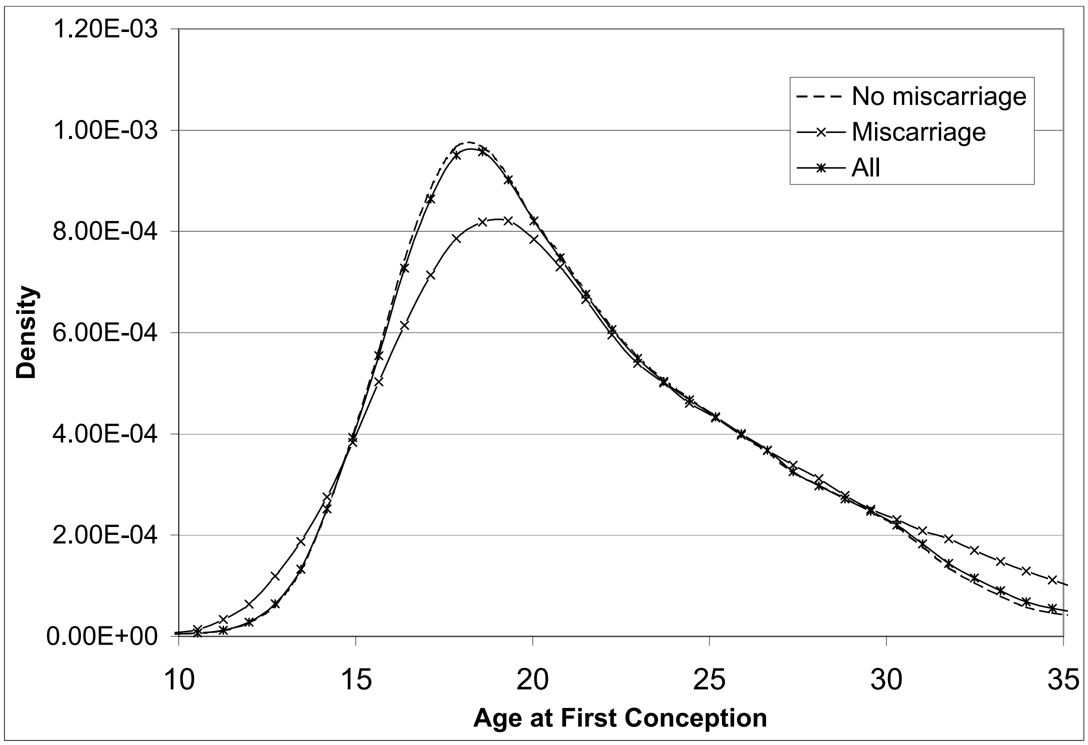


Figure 3.9: Kernel Density Estimates of Age at First Conception for Women who Have and who Have Not Experienced a Miscarriage



Appendix A

Definition of Variables for First Essay

The variables *age at wedding*, *age at interview*, and *age difference* between the spouses are self-explanatory.

Duration of the Marriage (continuous) Women were asked in which month their marriages started and ended. The end of a marriage is defined as the date when the spouses either separate or the marriage is officially ended by divorce or annulment. The official date of the divorce or the annulment can be later as the separation in which case the date of the separation marks the end of the marriage. The death of the husband censors the marriage at this date. All other marriages are censored at the date of the interview.

Premarital Cohabitation (dummy) Women in the NSFG 2002 were asked: "Some couples live together without being married. By living together, we mean having a sexual relationship while sharing the same usual address. Did you and (1st HUSBAND) live together before you got married?" The same question was asked with respect to their 2nd and 3rd husbands where applicable. A dummy with the values 1 and 0 was created depending on the answer.

Premarital Conception (dummy) Women were asked for all dates of conception for all their pregnancies. If this date was prior to the first marriage, the dummy for premarital conception takes on the value 1 for the first marriage. For higher order marriages, the conception has to be after the separation of the preceding marriage and before the current marriage to be considered a premarital conception.

Premarital Lifebirth (dummy) Similarly, women were asked for the dates of all their lifebirths. If this date was prior to the first marriage, the dummy for premarital conception takes on the value 1 for the first marriage. For higher order marriages, the conception has to be after the separation of the preceding marriage and before the current marriage to be considered a premarital conception.

Marital Lifebirth (dummy, continuous) For women with a lifebirth during their marriage a dummy was created. Furthermore, the time after the marriage date was created to record the time passed since marriage until the first lifebirth during marriage.

Educational achievement (dummy) Dummies for educational achievement (at the date of the interview) were created. Less than high school means that the woman has not completed 12 years of schooling. A dummy for high school indicates exactly 12 years of schooling while a dummy for more than high school indicates more years of schooling. It was unfortunately not possible to measure educational achievement at the date when the cohabitation began.

Race, Religion (dummy) Dummies for race and religion were created. Again, the dummy for religion are not time-varying but refer to the religion at the date of the interview.

Intact Family Background (dummy) A dummy for an intact family background was coded one if both parents (biological and adoptive) where present in the household until age 18 for the NSFG 1988 and 2002. For 1995, additionally there were distinct categories if both parents were present for biological and adoptive parents.

Appendix B

Proof of proposition.

Claim. $\varepsilon_{t+1}^{**}(3,2)$ is a strictly decreasing function in $f_t(3)$, and a strictly increasing function in $f_t(2)$.

Proof. The reservation value $\varepsilon_{t+1}^{**}(3,2)$ for an agent is implicitly defined as

$$V_t\left(3,2,\widehat{\theta}_{d_t},\varepsilon_{t+1}^{**}(3,2)\right)=V_t\left(2,2,\widehat{\theta}_{d_t},\varepsilon_{t+1}^{**}(3,2)\right) \tag{B.1}$$

$$\iff 0=f_t(3)-f_t(2) \tag{B.2}$$

$$+\beta\int_{-\infty}^{\infty}\left[\max_{m_{t+1}\in F(3)}V_{t+1}\left(m_{t+1},3,\widehat{\theta}_{d_{t+1}},\varepsilon_{t+1}\right)-\max_{m_{t+1}\in F(2)}V_{t+1}\left(m_{t+1},2,\widehat{\theta}_{d_{t+1}},\varepsilon_{t+1}\right)\right]$$

$$\times\phi\left(\varepsilon_{t+1}\middle|\widehat{\theta}_{d_t},\varepsilon_t\right)d\varepsilon_{t+1}$$

In the following all arguments in the value functions except for the indicator for cohabitation or marriage are dropped. Brien, Lillard, and Stern (1998) show that this is equal to

$$\begin{aligned} 0 &= f_t(3) - f_t(2) \\ &+\beta \int_{-\infty}^{\varepsilon_{t+1}^{**}(3,2)} \left[\max_{m_{t+1}\in F(3)} V_{t+1}(3) - \max_{m_{t+1}\in F(2)} V_{t+1}(2) \right] \phi\left(\varepsilon_{t+1} \middle| \widehat{\theta}_{d_t}, \varepsilon_t\right) d\varepsilon_{t+1} \end{aligned} \tag{B.3}$$

The derivative of the reservation value with respect to the flow utility is given by:

$$\frac{\partial \varepsilon_{t+1}^{**}(3,2)}{\partial f_t(3)} = -\frac{1}{\beta \left[A + \int_{-\infty}^{\varepsilon_{t+1}^{**}(3,2)} \frac{\partial \max_{m_{t+1}\in F(3)} V_{t+1}(3)}{\partial \varepsilon_{t+1}} - \frac{\partial \max_{m_{t+1}\in F(2)} V_{t+1}(2)}{\partial \varepsilon_{t+1}} \phi\left(\varepsilon_{t+1} \middle| \widehat{\theta}_{d_t}, \varepsilon_t\right) d\varepsilon_{t+1} \right]} \tag{B.4}$$

where $A = \max_{m_{t+1}\in F(3)} V_{t+1}\left(m_{t+1}, 3, \widehat{\theta}_{d_{t+1}}, \varepsilon_{t+1}^{**}\right) - \max_{m_{t+1}\in F(2)} V_{t+1}\left(m_{t+1}, 2, \widehat{\theta}_{d_{t+1}}, \varepsilon_{t+1}^{**}\right)$. The first expression A is equal to zero since at the reservation value the agent is just indifferent between getting married or staying in a cohabitation union. For this reason:

$$\begin{aligned} \frac{\partial \varepsilon_{t+1}^{**}(3,2)}{\partial f_t(3)} &= \\ -\frac{1}{+\beta \left[\int_{-\infty}^{\varepsilon_{t+1}^{**}(3,2)} \frac{\partial \max_{m_{t+1}\in F(3)} V_{t+1}(3)}{\delta \varepsilon_{t+1}} - \frac{\partial \max_{m_{t+1}\in F(2)} V_{t+1}(2)}{\delta \varepsilon_{t+1}} \phi\left(\varepsilon_{t+1} \middle| \widehat{\theta}_{d_t}, \varepsilon_t\right) d\varepsilon_{t+1} \right]} &= \\ -\frac{1}{\beta \left[\int_{\varepsilon_{t+1}^{*}(1,3)}^{\varepsilon_{t+1}^{**}(3,2)} \frac{\partial V_{t+1}(3)}{\partial \varepsilon_{t+1}} \phi\left(\varepsilon_{t+1} \middle| \widehat{\theta}_{d_t}, \varepsilon_t\right) d\varepsilon_{t+1} - \int_{\varepsilon_{t+1}^{*}(1,2)}^{\varepsilon_{t+1}^{**}(3,2)} \frac{\partial V_{t+1}(2)}{\partial \varepsilon_{t+1}} \phi\left(\varepsilon_{t+1} \middle| \widehat{\theta}_{d_t}, \varepsilon_t\right) d\varepsilon_{t+1} \right]} & \end{aligned} \tag{B.5}$$

They show also (see Brien, Lillard, and Stern 1998 and 2004) that $\varepsilon_{t+1}^{**}(3,2) > \varepsilon_{t+1}^{*}(1,2) > \varepsilon_{t+1}^{*}(1,3)$ and that $\frac{\partial V_t\left(3, m_{t-1}, \widehat{\theta}_{d_t}, \varepsilon_t\right)}{\partial \varepsilon_t} > \frac{\partial V_t\left(2, m_{t-1}, \widehat{\theta}_{d_t}, \varepsilon_t\right)}{\partial \varepsilon_t}$. For this reason the denominator is positive. It follows that

$$\frac{\partial \varepsilon_{t+1}^{**}(3,2)}{\partial f_t(3)} < 0 \tag{B.6}$$

Similar arguments can be applied to show that $\frac{\partial \varepsilon_{t+1}^{**}(3,2)}{\partial f_t(2)} > 0$. □

Bibliography

Abadie, Alberto, David Drukker, Jane Leber Herr, and Guido W. Imbens (2004): "Implementing Matching Estimators For Average Treatment Effects in Stata," *Stata Journal*, 4(3), 290–311.

Abadie, Alberto, and Guido W. Imbens (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74(1), 235–267.

Abrevaya, Jason (1999): "Leapfrog Estimation of a Fixed-Effects Model with Unknown Transformation of the Dependent Variable," *Journal of Econometrics*, 93, 203–228.

——— (2000): "Rank Estimation of a Generalized Fixed-Effect Regression Model," *Journal of Econometric*, 95, 1–23.

Aldrich, John H., and Forrest D. Nelson (1984): *Linear Probability, Logit, and Probit Models*. Beverly Hills, CA: Sage Publications.

Amato, Paul R., and Alan Booth (1995): "Changes in Gender Role Attitudes and Perceived Marital Quality," *American Sociological Review*, 60, 58–66.

Angrist, Joshua D., and Guido W. Imbens (1995): "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Society*, 90(430), 431–442.

Angrist, Joshua D., and Alan B. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings," *Quarterly Journal of Economics*, 106(4), 979–1014.

——— (1999): "Empirical Strategies in Labor Economics," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card. Amsterdam: Elsevier.

Becker, Gary S. (1973): "A Theory of Marriage," in *The Economics of the Family*, ed. by T. W. Schultz. Chicago: University of Chicago Press.

——— (1981): *A Treatise on the Family*. Cambridge, MA: Harvard University Press.

Becker, Gary S., Elisabeth M. Landes, and Robert T. Michael (1977): "An Economic Analysis of Marital Stability," *Journal of Political Economy*, 85(6), 1141–1188.

Bennett, Neil G., Ann Klimas Blanc, and David E. Bloom (1988): "Commitment and the Modern Union: Assessing the Link between Premarital Cohabitation and Subsequent Marital Stability," *American Sociological Review*, 53(1).

Bjorklund, Anders, and Robert A. Moffitt (1987): "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models," *Review of Economics and Statistics*, 69(1), 42–49.

Blake, Judith (1968): "Are Babies Consumer Durables? A Critique of the Economic Theory of Reproductive Motivation," *Population Studies*, 22(1), 5–25.

Booth, Alan, and David Johnson (1988): "Premarital Cohabitation and Marital Success," *Journal of Family*, 9(2), 255–272.

Brien, Michael J., Lee A. Lillard, and Steven Stern (2006): "Cohabitation, Marriage, and Divorce in a Model of Match Quality," *International Economic Review*, 47(2), 451–494.

Bronars, Stephen G., and Jeff Grogger (1993): "The Socioeconomic Consequences of Teenage Childbearing: Findings From a Natural Experiment," *Family Planning Perspectives*, 25(4), 156–161+174.

Bumpass, Larry L., and Hsien-Hen Lu (2000): "Trends in Cohabitation and Implications for Children's Family Contexts in the United States," *Population Studies*, 54.

Bumpass, Larry L., and James A. Sweet (1989): "National Estimates of Cohabitation," *Demography*, 26(4), 615–625.

Bumpass, Larry L., James A. Sweet, and Andrew J. Cherlin (1991): "The Role of Cohabitation in Declining Rates of Marriage," *Journal of Marriage and the Family*, 53(4), 913–927.

Cameron, Stephen V., and James J. Heckman (1993): "The Nonequivalence of High School Equivalents," *Journal of Labor Economics*, 11(1), 1–47.

Cherlin, Andrew J. (1991): *Marriage, Divorce, Remarriage*. Cambridge, MA: Harvard University Press.

——— (2004): "The Deinstitutionalization of American Marriage," *Journal of Marriage and the Family*, 66, 848–861.

Cox, David R. (1972): "Regression Models and Lifetables," *Journal of the Royal Statistical Society. Series B*, 34(2).

DeMaris, Alfred, and K. Vaninadha Rao (1992): "Premarital Cohabitation and Subsequent Marital Stability in the United States: A Reassessment," *Journal of Marriage and the Family*, 54(1).

for Health Statistics, National Center (1988, 1995, and 2002): "National Survey of Family Growth," .

Fraser, Alison M., John E. Brockert, and R.H. Ward (1995): "Association of Young Maternal Age with Adverse Reproductive Outcomes," *New England Journal of Medicine*, 332(17), 1113–1117.

Freedman, David S., Laura Kettel Khan, Mary K. Serdula, William H. Dietz, Sathanur R. Srinivasan, and Gerald S. Berenson (2002): "Relation of Age at Menarche to Race, Time Period, and Anthropometric Dimensions: The Bogalusa Heart Study," *Pediatrics*, 110(4), 1–7.

Gerfin, Michael, and Michael Lechner (2002): "A Microeconometric Evaluation of the Active Labour Market Policy in Switzerland," *Economic Journal*, 112, 854–893.

Geronimus, Arline T., and Sanders Korenman (1992): "The Socioeconomic Consequences of Teen Childbearing Reconsidered," *The Quarterly Journal of Economics*, 107(4), 1187–1214.

Hansen, Lars Peter (1982): "Large Sample Properties of Generalized Methods of Moments," *Econometrica*, 50(4), 1029–1054.

Heckman, James J. (1978): "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46(5), 931–960.

——— (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–161.

Heckman, James J., and Richard Robb (1985): "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data*, ed. by J. J. Heckman, and B. Singer. Cambridge, UK: Cambridge University Press.

Heckman, James J., Sergio Urzua, and Edward Vytlacil (2006): "Understanding Instrumental Variables in Models with Essential Heterogeneity," *Review of Economics and Statistics*, 88(3), 389–432.

Heckman, James J., and Edward Vytlacil (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluations," *Econometrica*, 73(3), 669–738.

Hotz, V. Joseph, Susan Williams McElroy, and Seth G. Sanders (1997): "The Costs and Consequences of Teenage Childbearing for Mothers and Government," in *Kids Having Kids*, ed. by R. A. Maynard. Washington D.C.: Urban Institute Press.

——— (1999): "Teenage Childbearing and its Life Cycle Consequences. Exploiting a Natural Experiment," NBER Working Paper 7397.

Imbens, Guido W., and Joshua D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

Khan, Shakeeb, and Elie Tamer (2006): "Partial Rank Estimation of Duration Models with Generalized Forms of Censoring," *Journal of Econometrics*, forthcoming.

Lancaster, Tony (1979): "Econometric Methods for the Duration of Unemployment," *Econometrica*, 47(4), 939–956.

Lee, Lung-Fei (1978): "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables," *International Economic Review*, 19(2), 415–433.

Lee, Sokbae (2003): "Estimating Panel Data Duration Models with Censored Data," Mimeo, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Lichter, Daniel T., Diane K. MacLaughlin, George Kephart, and David J. Landry (1992): "Race and the Retreat from Marriage: A Shortage of Marriageable Men?," *American Sociological Review*, 57(6), 781–799.

Liefbroer, Aart C., and Edith Dourleijn (2006): "Unmarried Cohabitation and Union Stability: Testing the Role of Diffusion Using Data from 16 European Countries," *Demography*, 43(2), 203–221.

Lillard, Lee A., Michael J. Brien, and Linda J. Waite (1995): "Premarital Cohabitation and Subsequent Marital Dissolution: A Matter of Self-Selection?," *Demography*, 32(3), 437–457.

Manning, Wendy D., and Pamela J. Smock (2005): "Measuring and Modeling Cohabitation: New Perspectives from Qualitative Data," *Journal of Marriage and the Family*, 67, 989–1002.

Moffitt, Robert A. (2000): "Female Wages, Male Wages, and the Economic Model of Marriage: The Basic Evidence," in *The Ties that Bind. Perspectives on Marriage and Cohabitation*, ed. by L. J. Waite. New York: de Gruyter.

——— (2005): "Remarks on the Analysis of Causal Relationships in Population Research," *Demography*, 42(1), 91–108.

——— (2007): "Estimating Average and Marginal Treatment Effects in Heterogenous Populations," Mimeo. Johns Hopkins University.

Moffitt, Robert A., Robert Reville, and Anne E. Winkler (1998): "Beyond Single Mothers: Cohabitation and Marriage in the ADFC Program," *Demography*, 35(3), 259–278.

Ribar, David C. (1994): "Teenage Fertility and High School Completion," *Review of Economics and Statistics*, 76(3), 413–424.

Roy, A.D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers, New Series*, 3(2), 135–146.

Schoen, Robert (1992): "First Unions and the Stability of First Marriages," *Journal of Marriage and the Family*, 54(2), 281–284.

Song, Xue (2001): "Labor supply and fertility behaviors in cohabitation and marriage.," Ph.D. thesis, Johns Hopkins University.

Staiger, Douglas, and James H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65(3), 557–586.

Stock, James H., and Motohiro Yogo (2002): "Testing for Weak Instruments in Linear IV Regressions," NBER Technical Working Paper, T284.

Svarer, Michael (2004): "Is Your Love in Vain? Another Look at Premarital Cohabitation and Divorce," *Journal of Human Resources*, 39(2), 523–535.

Teachman, Jay D. (2002): "Premarital Sex, Premarital Cohabitation, and the Risk of Subsequent Marital Dissolution among Women," *Journal of Marriage and the Family*, 65, 444–455.

——— (2003): "Stability across Cohorts in Divorce Risk Factors," *Demography*, 39(2), 331–351.

Teachman, Jay D., and Karen A. Polonko (1990): "Cohabitation and Marital Stability in the United States," *Social Forces*, 60(1), 207–220.

Van den Berg, Gerald J. (2001): "Duration Models: Specification, Identification and

Multiple Durations," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, vol. 5. Amsterdam: Elsevier.

Curriculum Vitae

Steffen Reinhold was born in Marburg, Germany on October 27th, 1976. He graduated as Diplom-Volkswirt (equivalent to an M.A. in Business Economics) from the Freie Universität Berlin, Germany in August 2003. He spent one year as an exchange student at the Johns Hopkins University in 2001-2002. In 2003, he entered the Johns Hopkins University in Baltimore where he completed a Master of Arts in Economics in 2006 and a Doctor of Philosophy in 2007. In September 2007, he will start working at the Department of Economics at the University of Mannheim, Germany.